



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 119737

TO: Michael Borin
Location: rem 2a55
Art Unit: 1631
Thursday, April 22, 2004

2170

Case Serial Number: 10/078090

From: Toby Port
Location: Biotech-Chem Library
Remsen 1A59
Phone: 571-272-2523

toby.port@uspto.gov

Search Notes

Dear Examiner Borin,

Here are the results of your search.
Please feel free to contact me if you have any questions.

Toby Port

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 21, 2004, 21:08:41 ; Search time 703.997 Seconds
(without alignments)
10740.063 Million cell updates/sec

Title: US-10-078-090-48

Perfect score: 1677

Sequence: 1 gagttgcgcgtgccaaaggc.....aagcgagaccaggggggaga 1677

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2907579 seqs, 2254313464 residues

Total number of hits satisfying chosen parameters: 5815158

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US05_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
- 17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 18: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
- 19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1677	100.0	1677	15	US-10-078-090-48
2	647.2	38.6	667	15	Sequence 48, Appl
3	554.8	33.1	2964	13	Sequence 47, Appl
4	420.2	25.1	1718	16	Sequence 536, App
5	240.8	14.4	509	15	Sequence 782, App
6	195	11.6	195	15	Sequence 3728, App
7	107.6	6.4	2822	16	Sequence 17428, A
8	76.6	4.6	553	13	Sequence 406, App
9	76.6	4.6	553	13	Sequence 285026, A
10	76.6	4.6	553	16	Sequence 285026, A
11	76.6	4.6	553	16	Sequence 285026, A
12	76.2	4.5	553	13	Sequence 285027, A
13	76.2	4.5	553	16	Sequence 285027, A
14	76.2	4.5	113633	13	Sequence 19, Appl

ALIGNMENTS

RESULT 1
US-10-078-090-48
Sequence 48, Application US/10078090
Publication No. US20030044815A1

- GENERAL INFORMATION:
APPLICANT: Salceda, Susana
APPLICANT: Hu, Ping
APPLICANT: Recipon, Hervé
APPLICANT: Karra, Kalpana
APPLICANT: Cafferkey, Robert
APPLICANT: Sun, Yongming
APPLICANT: Liu, Chenghua
TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Pr

FILE REFERENCE: DEX-0312
CURRENT APPLICATION NUMBER: US/10/078,090
CURRENT FILING DATE: 2002-02-14
PRIOR APPLICATION NUMBER: 60/268,999
PRIOR FILING DATE: 2001-02-15
NUMBER OF SEQ ID NOS: 210
SOFTWARE: PatentIn version 3.1
SEQ ID NO 48
LENGTH: 1677
TYPE: DNA
ORGANISM: Homo sapien
US-10-078-090-48

Query Match 100.0%; Score 1677; DB 15; Length 1677;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1677; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAGTTGCGGCGTGCAGGCCCCAGAGGGCTTCAGATCCGTGGGGGCTCGGAG 60

Db 1 GAGTTGCGGCGTGCAGGCCCCAGAGGGCTTCAGATCCGTGGGGGCTCGGAG 60

Qy 61 CACGGCGTGGGATCTACGTGCTCTGTGGGAACAGAGGCTCTAGCTCAGAGGAAGA 120

|||||

61 CACGGCTGGGCATCTACGTGCTCTCTGGTGAACCCAGGCTCTCTAGCTGAGAGGAAGGA 120
121 CTGGGGTGGGACCAAGATCTCGCGCTCAACGACAAATCCCTGGCCCGGGTGACCCAC 180
121 CTGGGGTGGGACCAAGATCTCGCGCTCAACGACAAATCCCTGGCCCGGGTGACCCAC 180
181 GCGAGGCGCGTCAAGAGGCTCTGAAGGGCTCCAAGAGCTGGTGTCTGTGTACTCAGCA 240
181 GCGAGGCGCGTCAAGAGGCTCTGAAGGGCTCCAAGAGCTGGTGTCTGTGTACTCAGCA 240
241 GGGGGCATCTGGGGGTGCTGACCAACACATCTACACCTGGGTGGACCCGCGAGGGC 300
241 GGGGGCATCTGGGGGTGCTGACCAACACATCTACACCTGGGTGGACCCGCGAGGGC 300
301 GCGAGCATCTCCCAACCTCGGGGCTGCCCGAGCCCAACGCTGGTGTCTGTACTGAGCAGCAG 360
301 GCGAGCATCTCCCAACCTCGGGGCTGCCCGAGCCCAACGCTGGTGTCTGTACTGAGCAGCAG 360
361 GAGGTGACCGGAGAGACCTGCACTCTGCAAGAGGGGATGAGAAAAGGTGAAC 420
361 GAGGTGACCGGAGAGACCTGCACTCTGCAAGAGGGGATGAGAAAAGGTGAAC 420
421 CTGGTGTGGGGACGGCGGCTCCCTGGGCTCACGATCCGTGGGGAGCTGAGTACGGC 480
421 CTGGTGTGGGGACGGCGGCTCCCTGGGCTCACGATCCGTGGGGAGCTGAGTACGGC 480
481 CTGGCATTTACATCACTGGCGTGAACCGAGCTCTGAAGCAGAGGAGCGGCTCAAG 540
481 CTGGCATTTACATCACTGGCGTGAACCGAGCTCTGAAGCAGAGGAGCGGCTCAAG 540
541 GTTGGGACCAATTTAGAGTGAATGGGGAGCTTTCTCAACATCCTACAGAGCAG 600
541 GTTGGGACCAATTTAGAGTGAATGGGGAGCTTTCTCAACATCCTACAGAGCAG 600
601 GCTGTACGGCTGCTTAAGTCACTCGGCACTCTGACAGTGAAGAGAGCTGGGAGG 660
601 GCTGTACGGCTGCTTAAGTCACTCGGCACTCTGACAGTGAAGAGAGCTGGGAGG 660
661 CTGCCCCATGCGGACCACTGTGACGAGACCAAGTGGATCGCCAGTTCGCCGATCAGG 720
661 CTGCCCCATGCGGACCACTGTGACGAGACCAAGTGGATCGCCAGTTCGCCGATCAGG 720
721 GAGACCATGGCAACTCGGACGGGTGTGGCCACTGTGCTGGTCCAAATCTCCAGACCCCA 780
721 GAGACCATGGCAACTCGGACGGGTGTGGCCACTGTGCTGGTCCAAATCTCCAGACCCCA 780
781 GGGCCATTTGAAAGCGAGTGAATGCTGCTCCCATCCCTCCAGCGCTGGCTCTCT 840
781 GGGCCATTTGAAAGCGAGTGAATGCTGCTCCCATCCCTCCAGCGCTGGCTCTCT 840
841 CTGAGCTGCACTGCGGACCAACAGGCGCTCCATTTGGCAGGACATGACCTGGGCATCC 900
841 CTGAGCTGCACTGCGGACCAACAGGCGCTCCATTTGGCAGGACATGACCTGGGCATCC 900
901 CTCTCTCTTGGGCTCAGTTTCCCATGGAAGCTGAAATACACCATCCCACTGCTCTC 960
901 CTCTCTCTTGGGCTCAGTTTCCCATGGAAGCTGAAATACACCATCCCACTGCTCTC 960
961 ATTCTTTATTGTGCCCCAAATTAATCTTAATCTATAGACCTTATGTTGCTTCATCCAA 1020
961 ATTCTTTATTGTGCCCCAAATTAATCTTAATCTATAGACCTTATGTTGCTTCATCCAA 1020
1021 AAAGTGGGACCAATACCCCTGCGCTCATCCAGATCTGTGCAGATGAAGAGAGGGAGG 1080
1021 AAAGTGGGACCAATACCCCTGCGCTCATCCAGATCTGTGCAGATGAAGAGAGGGAGG 1080
1081 AGAGGGAAGAGAGAGATGTTTGGGGTGTATTGGCCAGAGGCCACAGGCTGGATCC 1140
1081 AGAGGGAAGAGAGAGATGTTTGGGGTGTATTGGCCAGAGGCCACAGGCTGGATCC 1140
1141 ATGAAGAAATCTGGGTGAGAGGGTCTTAAAGTCAATAAATGAGATCCAGTTCAGGTGG 1200
1141 ATGAAGAAATCTGGGTGAGAGGGTCTTAAAGTCAATAAATGAGATCCAGTTCAGGTGG 1200

RESULT 2

US-10-078-090-47
; Sequence 47, Application US/10078090
; Publication No. US20030044815A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Hu, Ping
; APPLICANT: Recipon, Herve
; APPLICANT: Karra, Kalpana
; APPLICANT: Caferkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Pr
; FILE REFERENCE: DEX-0312
; CURRENT APPLICATION NUMBER: US/10/078,090
; CURRENT FILING DATE: 2002-02-14
; PRIOR APPLICATION NUMBER: 60/268,999
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 210
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 47
; LENGTH: 667
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-078-090-47

Query Match 38.6%; Score 647.2; DB 15; Length 667;

Best Local Similarity 99.5%; Pred. No. 2e-177;

Matches 649; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1026 GGGGACCAATAACCTTGGCTTATTTGGCCAGAGGCCACAGGCTGGATCCCATGAA 1145
Db 16 GAGGTCCATAACCTTGGCTTATTTGGCCAGAGGCCACAGGCTGGATCCCATGAA 135
QY 1086 GAAAGAGAGAGATGCTTTGGGGTGTATTGGCCAGAGGCCACAGGCTGGATCCCATGAA 1145
Db 76 GAAAGAGAGAGATGCTTTGGGGTGTATTGGCCAGAGGCCACAGGCTGGATCCCATGAA 135


```
; APPLICANT: SEKI, NAHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 782
; LENGTH: 1718
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-094-749-782

Query Match      25.1%; Score 420.2; DB 16; Length 1718;
Best Local Similarity 95.0%; Pred. No. 4.1e-111;
Matches 434; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 4 TTGCGGCGTGCACAGGCCACAGGGCTTGGGCTTCAGCATCCGTGGGGGCTCGAGCAC 63
Db 1076 TTGCGGCGTGCACAGGCCACAGGGCTTGGGCTTCAGCATCCGTGGGGGCTCGAGCAC 1135

QY 64 GCGTGGGCGATCTACGTGTCTCTGTGGAACAGAGCTCTCTAGCTGAGAAAGGAGCTG 123
Db 1136 GCGTGGGCGATCTACGTGTCTCTGTGGAACAGAGCTCTCTAGCTGAGAAAGGAGCTG 1195

QY 124 CGGTCGGGGACAGATTCTGCGGTCAACGACAAATCCCTGGCCCGGGTACCCACGCG 183
Db 1196 CGGTCGGGGACAGATTCTGCGGTCAACGACAAATCCCTGGCCCGGGTACCCACGCG 1255

QY 184 GAGGCGCTCAAGCTCTGAAGGCTCCAAAGAGCTGTCTGTCTGTACTCAGCAGGG 243
Db 1256 GAGGCGCTCAAGCTCTGAAGGCTCCAAAGAGCTGTCTGTCTGTACTCAGCAGGG 1315

QY 244 CGCATCCCTGGGGGTACTGTCAACACCATCTACACTGGGTGGACCCGAGGCGCG 303
Db 1316 CGCATCTCTGGGGGTACTGTCAACACCATCTACACTGGGTGGACCCGAGGCGCG 1375

QY 304 AGCATCTCCACCTCGGGCTGCCAGCCCGGCTGCGGCTGAGGCGAGCAGGAG 363
Db 1376 AGCATCTCCACCTCGGGCTGCCAGCCCGGCTGCGGCTGAGGCGAGCAGGAG 1435

QY 364 GGTGACCGGAGGAGCACCCTGCACCTCTCTGCAAGGAGGGGATGAGAAAAGGTGAACCTG 423
Db 1436 GGTGACCGGAGGAGCACCCTGCACCTCTCTGCAAGGAGGGGATGAGAAAAGGTGAGTGGG 1495

QY 424 GTCTCGGGGAGCGGCGGCTCCCTGGGCTTCACGATCC 460
Db 1496 GTGGGAAAGGAGGCGGCGCTTCACGACACCGTATTC 1532

RESULT 5
US-10-029-386-3728
; Sequence 3728, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

; SEQ ID NO 17428
; LENGTH: 195
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL138895.4
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.63
; OTHER INFORMATION: EST HUMAN HIT: B8890168.1, EVALUE 1.00e-105
; OTHER INFORMATION: NT HIT: GI14735271, EVALUE 1.00e-106
; OTHER INFORMATION: SWISSPROT HIT: P35428, EVALUE 3.70e-01
US-10-029-386-17428
```

```
; SEQ ID NO 3728
; LENGTH: 509
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL138895.4
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.63
; OTHER INFORMATION: NT HIT: GI14735271, EVALUE 1.00e-123
; OTHER INFORMATION: EST HUMAN HIT: BP969269.1, EVALUE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P54817, EVALUE 3.70e-02
US-10-029-386-3728

Query Match      14.4%; Score 240.8; DB 15; Length 509;
Best Local Similarity 77.1%; Pred. No. 3.2e-59;
Matches 293; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 81 GTCTCTGTTGGAACACAGGCTCTCTAGCTGAGAAAGGAGGACTGCGGGTCGGGACCCAGAT 140
Db 72 GCCTCTTCTTCCATGTGGGGCTCTCTGAGAAACCAAGGCTTGGGCCACGGGAGGGTGT 131

QY 141 TCTGCGCGTCAACGACAAATCCCTGGCCCGGGTGACCCACCGGAGGCGGTCAAGGCTCT 200
Db 132 CTCTCCACACCCCGGACCTCCCTCTCTAACTTTGTATCCCTCTCTCTAGGCTCT 191

QY 201 GAAGGGCTCCAAAGCTGTGTGTCTGTACTCAGCAGGCGCATCCCTGGGGGCTA 260
Db 192 GAAGGGCTCCAAAGCTGTGTGTCTGTACTCAGCAGGCGCATCCCTGGGGGCTA 251

QY 261 CGTCACCAACCATCTACACCTGGGTGACCCGACCGGCGCGAGCATTTCCCCACCCCTC 320
Db 252 CGTCACCAACCATCTACACCTGGGTGACCCGACCGGCGCGAGCATTTCCCCACCCCTC 311

QY 321 GGGCGCTGCCACCCACCGGTGTCCTGAGGACGAGGAGGTGACCGGAGGAGCAC 380
Db 312 GGGCGCTGCCACCCACCGGTGTCCTGAGGACGAGGAGGTGACCGGAGGAGCAC 371

QY 381 CCTGCACCTCTCTCAAGGAGGGGATGAGAAAAGGTGAACCTGTGTCTGGGGACCGCGG 440
Db 372 CCTGCACCTCTCTCAAGGAGGGGATGAGAAAAGGTGAGTGGGGTGGGAAAGGAGGCCA 431

QY 441 GTCTCTGGGCTCACGATCC 460
Db 432 GCCTCTCAGACACCGTATTC 451

RESULT 6
US-10-029-386-17428
; Sequence 17428, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

; SEQ ID NO 17428
; LENGTH: 195
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL138895.4
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.63
; OTHER INFORMATION: EST HUMAN HIT: B8890168.1, EVALUE 1.00e-105
; OTHER INFORMATION: NT HIT: GI14735271, EVALUE 1.00e-106
; OTHER INFORMATION: SWISSPROT HIT: P35428, EVALUE 3.70e-01
US-10-029-386-17428

Query Match      11.6%; Score 195; DB 15; Length 195;
```

Query March	6.4%;	Score 107.6;	DB 16;	Length 2822;
Best Local Similarity	49.9%;	Pred. No. 5.3e-20;		
Matches 361;	Conservative 0;	Mismatches 324;	Indels 39;	Gaps 2
QY	25	GAGGGCTTTGGG	TTCCAGCATCCG	TGGGGCTCGGAGCACGGCGTGGGCATCTACCTGTCT 84
DB	348	GGGAGGCTGGG	CTTCAGCGTGG	CGGGGGCTCAGAGCATGGCTGGGCATCTTCGTCTCAGC 407
QY	85	CTGGTGGACACAGG	CTCTACTCAGAGGAAG	AGACTTGGCGGTGGGGACCCAGATTTCTG 144
DB	408	AAAGTGGAGGAAGG	CAGCATGTCAGAGCGGG	TGCGCTTGGCTGGGGACAAATCAGC 467
QY	145	CGCGTCAACGACAA	ATCCCTGGCCCGGGTG	ATCCACCGGAGCGCGTCAAGGCTCTGAAG 204
DB	468	GAGGTGAATGGG	CTGAGCCTCGAGAG	CACCATGGGTAGCGCCGTAAAGGTGCTGACC 527
QY	205	GGCTCCAGAAAGCT	GGTGCTGTCTGTGTACT	CAGCAGGGCGCATCCCTGGGGGCTACGCTC 264
DB	528	AGCAGCAGCGGCT	GCACATGATGTTT	CGGCGCATGGGCCGTGTCCGGGCATCAAGTTC 587
QY	265	ACCAACCAATCTA	CACCTTGGGTGGAC	CCGCGAGGGCCGACAGCATCTCCCCACCTTCGGGC 324
DB	588	TCCAAGGAAGACCA	CGTGGGTGGATGTGGT	GAATCGCGCGCTGTTATGAGGAAGATGC 647

```

RESULT 8
US-10-027-632-285026
; Sequence 285026, Application US/10027632
; Publication NO. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,353
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 285026
; LENGTH: 553
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-285026

```


[illegible]

```

RESULT 12
US-10-027-632-285027
; Sequence 285027, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 285027
; LENGTH: 553
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-285027

```

	Query Match	4.5%;	Score 76.2;	DB 13;	Length 553;
	Best Local Similarity	63.7%;	Pred. No. 2.7e-11;		
	Matches 114;	Conservative 1;	Mismatches 64;	Indels 0;	Gaps 0;
Qy	1299	TCTGTCGTCTGTCTGTCTCTCTGTCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG			1358
Dd	343	TCTCTCTCTCTTCTCTCTTTCTGTGTGTGTGGGGGTGTGTGTGTGTGTGTG			402
Qy	1359	TGTGTGTGTGTGTCTCTCAGCTCTTTCAATCCCTCATATTACATAGTAGTATAATAATAA			1418
Dd	403	TGTGTGTGTGTCTCTCTCTCTCTGTCTGAAGGGGAATTCGAAAAATTGCTGCTGAATAA			462
Qy	1419	ATATTAGAGAGATACACAGAAAATATATPAGAGAAGATAACAGTGTTCTCTTAATAAAAAA			1477
Dd	463	TTTTTCTTGAAAAACCAGGTGTGATATGATGATATATTAATCTTCTCTATGAACAGA			521

```

RESULT 13
US-10-027-632-285027
; Sequence 285027, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129

```

```

/ CURRENT APPLICATION NUMBER: US/10/027,632
/ CURRENT FILING DATE: 2002-04-30
/ PRIOR APPLICATION NUMBER: US 60/218,006
/ PRIOR FILING DATE: 2000-07-12
/ PRIOR APPLICATION NUMBER: US 60/198,676
/ PRIOR FILING DATE: 2000-04-20
/ PRIOR APPLICATION NUMBER: US 60/193,483
/ PRIOR FILING DATE: 2000-03-29
/ PRIOR APPLICATION NUMBER: US 60/185,218
/ PRIOR FILING DATE: 2000-02-24
/ PRIOR APPLICATION NUMBER: US 60/167,363
/ PRIOR FILING DATE: 1999-11-23
/ PRIOR APPLICATION NUMBER: US 60/156,358
/ PRIOR FILING DATE: 1999-09-28
/ PRIOR APPLICATION NUMBER: US 60/146,002
/ PRIOR FILING DATE: 1999-08-09
/ NUMBER OF SEQ ID NOS: 325720
/ SOFTWARE: PASCSEQ for Windows Version 4.0
/ SEQ ID NO 285027
/ LENGTH: 553
/ TYPE: DNA
/ ORGANISM: Human
/ US-10-027-632-285027

```

[illegible]

```

RESULT 14
US-10-087-192-19
; Sequence 19, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AN
; FILE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 113633
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURES:
; NAME/KEY: misc_feature
; LOCATION: (1)..(113633)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-19

```

Query Match 4.5%; Score 76.2; DB 13; Length 113633;
Best Local Similarity 96.3%; Pred. No. 6.9e-10;
Matches 78; Conservative 0; Mismatches 3; Indels 0;

Search completed: April 22, 2004, 01:59:54
Job time : 707.997 secs

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

OM nucleic - nucleic search, using sw model

Run on: April 21, 2004, 19:28:36 ; Search time 2576.09 Seconds (without alignments)
11222.365 Million cell updates/sec

Title: US-10-078-090-47

Perfect score: 667

Sequence: 1 gcgtggcggggcccgaggt.....aagcgagaccaggggggaga 667

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.ste.*
12: gb.sv.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.ste.*
28: em.un.*
29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rod.*
36: em.htg.man.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

RESULT 1
ALI38895/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE

ALI38895
Human DNA sequence from clone Rp11-9M16 on chromosome 9, complete
sequence.
ALI38895
ALI38895.16 GI:14575083
HTG
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 132292)
Skuce,C.
Direct Submission

ALIGNMENTS

ALI38895 132292 bp DNA linear PRI 27-JUN-2001
Human DNA sequence from clone Rp11-9M16 on chromosome 9, complete
sequence.
ALI38895
ALI38895.16 GI:14575083
HTG
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 132292)
Skuce,C.
Direct Submission

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	346.2	51.9	132292	9	ALI38895 Human DNA
C 2	339.2	50.9	171827	2	AF336382 Homo sapi
C 3	89.6	13.4	2093	5	BC051776 Danio rer
C 4	87.8	13.2	139973	2	AC135372 Rattus no
C 5	86.6	13.2	214625	10	AC116582 Mus muscu
C 6	86.6	13.0	211588	2	EX537348 Danio rer
C 7	86.2	12.9	46643	9	AL732512 Mouse DNA
C 8	85.2	12.8	166893	9	AP002987 Homo sapi
C 9	85.2	12.8	179355	10	AC115631 Mus muscu
C 10	85.2	12.8	203271	2	EX322566 Danio rer
C 11	85	12.7	215778	2	EX530074 Danio rer
C 12	85	12.7	215778	2	EX530074 Danio rer
C 13	84.8	12.7	135904	2	AC139157 Mus muscu
C 14	84.8	12.7	213673	10	AC130218 Mus muscu
C 15	84.6	12.7	117886	2	AC139385 Rattus no
C 16	84.6	12.7	160718	2	EX664714 Danio rer
C 17	84.6	12.7	202878	2	EX511101 Danio rer
C 18	84.6	12.7	228962	2	AC131587 Mus muscu
C 19	84.6	12.7	229574	2	AC123856 Mus muscu
C 20	84.4	12.7	181884	2	AC121089 Mus muscu
C 21	84.4	12.7	205170	2	AC113482 Mus muscu
C 22	84.2	12.6	137393	2	EX545852 Danio rer
C 23	84	12.6	169095	5	EX08685 Zebrafish
C 24	84	12.6	230877	5	EX248500 Zebrafish
C 25	84	12.6	237343	2	EX569792 Danio rer
C 26	83.8	12.6	165993	5	AL772362 Zebrafish
C 27	83.6	12.5	142803	2	AC013771 Homo sapi
C 28	83.6	12.5	169086	2	AC015832 Homo sapi
C 29	83.6	12.5	169821	2	AC019098 Homo sapi
C 30	83.6	12.5	198495	9	AP003441 Homo sapi
C 31	83.6	12.5	247113	2	AC107008 Rattus no
C 32	83.6	12.5	256513	2	AC155517 Rattus no
C 33	83.4	12.5	130176	2	EX005295 Danio rer
C 34	83.4	12.5	171807	2	EX000463 Danio rer
C 35	83.4	12.5	183326	2	EX571886 Danio rer
C 36	83.2	12.5	263668	2	AC136184 Rattus no
C 37	83.2	12.5	323010	2	AC111643 Rattus no
C 38	83	12.4	127560	5	EX005077 Zebrafish
C 39	83	12.4	179186	10	AC125396 Mus muscu
C 40	83	12.4	196887	2	AC114586 Mus muscu
C 41	83	12.4	200207	2	EX323062 Danio rer
C 42	83	12.4	201642	10	AL672174 Mouse DNA
C 43	82.8	12.4	216853	2	EX284112 Danio rer
C 44	82.6	12.4	110000	2	EX572080 Danio rer
C 45	82.6	12.4	110000	2	Continuation (2 of

JOURNAL Submitted (27-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 COMMENT requests: clonerequest@sanger.ac.uk
 On Jun 28, 2001 this sequence version replaced gi:14456172.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; SW: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr-9>
 RP11-9M16 is from the library RP11-11.1 constructed by the group of Pieter de Jong. For further details see
<http://www.choxi.org/bacpac/home.htm>
 VECTOR: pBACe3.6
 IMPORTANT: This sequence is not the entire insert of clone RP11-9M16 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
 The true right end of clone RP11-9M16 is at 132292 in this sequence. The true left end of clone RP11-402G3 is at 124695 in this sequence. The true right end of clone RP11-8211 is at 2000 in this sequence.
 Location/Qualifiers
 1..132292
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="9"
 /clone="RP11-9M16"
 /clone_lib="RP11-11.1"
 1..32
 /note="AluY repeat: matches 1..32 of consensus"
 74..301
 /note="L1MC1 repeat: matches 5718..5955 of consensus"
 399..696
 /note="AluSq repeat: matches 1..298 of consensus"
 1186..1273
 /note="MIR repeat: matches 49..142 of consensus"
 2271..2435
 /note="MER63A repeat: matches 46..206 of consensus"
 2467..2553
 /note="MIR repeat: matches 79..165 of consensus"
 2587..2713
 /note="L2 repeat: matches 2622..2740 of consensus"
 2911..3207
 /note="AluSx repeat: matches 5..301 of consensus"
 3214..3262
 /note="MIR repeat: matches 115..166 of consensus"
 4193..4262
 /note="MIR repeat: matches 60..130 of consensus"
 4368..5308
 /note="CpG island"
 /evidence="not_experimental"
 5299..5605
 /note="AluSx repeat: matches 12..312 of consensus"
 5686..5787
 /note="MIR repeat: matches 141..262 of consensus"
 6714..6799

/note="MIR repeat: matches 20..114 of consensus"
 6995..7437
 /note="MT1D repeat: matches 3..503 of consensus"
 7440..7502
 /note="MIR repeat: matches 82..144 of consensus"
 8034..8179
 /note="LFR16A repeat: matches 200..354 of consensus"
 8160..8271
 /note="LFR16C repeat: matches 255..368 of consensus"
 10037..10131
 /note="MIR repeat: matches 14..123 of consensus"
 10604..10711
 /note="L2 repeat: matches 2405..2521 of consensus"
 11151..11327
 /note="MIR repeat: matches 18..205 of consensus"
 11579..11883
 /note="AluSx repeat: matches 1..304 of consensus"
 11886..12024
 /note="MIR repeat: matches 97..243 of consensus"
 12598..12717
 /note="MIR repeat: matches 52..181 of consensus"
 13839..14002
 /note="MIR repeat: matches 80..258 of consensus"
 14071..14205
 /note="MIR repeat: matches 92..226 of consensus"
 15216..15653
 /note="MT2CB repeat: matches 1..492 of consensus"
 15702..15995
 /note="L2 repeat: matches 2372..2688 of consensus"
 16113..16847
 /note="L2 repeat: matches 1940..2707 of consensus"
 16875..16922
 /note="L2 copies 4 mer atga 81% conserved"
 17482..17573
 /note="MIR repeat: matches 125..230 of consensus"
 18163..18379
 /note="MIR repeat: matches 23..241 of consensus"
 18972..19080
 /note="MIR repeat: matches 34..144 of consensus"
 20542..20994
 /note="Charlie4 repeat: matches 26..506 of consensus"
 21421..21521
 /note="L2 repeat: matches 2045..2158 of consensus"
 21523..21758
 /note="AluJo repeat: matches 1..232 of consensus"
 22863..23243
 /note="MSTA repeat: matches 1..426 of consensus"
 23319..23495
 /note="MIR repeat: matches 5..192 of consensus"
 24438..24542
 /note="MIR repeat: matches 28..140 of consensus"
 24993..25526
 /note="L2 repeat: matches 2099..2709 of consensus"
 25630..25867
 /note="MIR repeat: matches 16..262 of consensus"
 25889..26233
 /note="L2 repeat: matches 2363..2674 of consensus"
 26257..26381
 /note="MIR repeat: matches 28..153 of consensus"
 26919..27000
 /note="MIR repeat: matches 64..147 of consensus"
 27292..27435
 /note="MIR repeat: matches 102..254 of consensus"
 29060..29358
 /note="AluSx repeat: matches 1..298 of consensus"
 29523..29664
 /note="L2 repeat: matches 2593..2734 of consensus"
 29669..29777
 /note="MIR repeat: matches 17..140 of consensus"
 30134..30243
 /note="L2 repeat: matches 2167..2289 of consensus"
 30460..30627
 /note="MER5B repeat: matches 1..173 of consensus"

Query Match	51.9%;	Score	346.2;	DB 9;	Length	132292;			
Best Local Similarity	89.6%;	Pred. No.	9.4e-80;						
Matches	372;	Conservative	0;	Mismatches	43;	Indels	0;	Gaps	0;

QY	16	GAGGTCCATAACCTCGCCCTCATCCAGATCTGTGCAGATGAAGAGAGGGAGGAGAGG	75
DB	32436	GGGGACCATAAACCTCGCCCTCATCCAGATCTGTGCAGATGAAGAGAGGGAGGAGAGG	32377
QY	76	GAAGAGAGAGATGCTTTGGGGTGATATTGGCCAGAGGCCACAGCTGGATGCCATGAA	135
DB	32376	GAAGAGAGAGATGCTTTGGGGTGATATTGGCCAGAGGCCACAGCTGGATGCCATGAA	32317
QY	136	GAATCTGGGTGAGAGGGTCTTAAAGTCATAAAGTCAGATCCAGTTGCCAGTGGCTGCA	195
DB	32316	GAATCTGGGTGAGAGGGTCTTAAAGTCATAAAGTCAGATCCAGTTGCCAGTGGCTGCA	32257
QY	196	TAGTTGCCAACAGTGTAAATGTGTCACTTTTGATCTTCATCAGAAATCTCAGCCTGGTG	255
DB	32256	TAGTTGCCAACAGTGTAAATGTGTCACTTTTGATCTTCATCAGAAATCTCAGCCTGGTG	32197
QY	256	CCACCTGGCCAAATACACTGCAGAGCATGTGTGTGTGTGTGTGTGTGTGTCTCTGT	315
DB	32196	CCACCTGGCCAAATACACTGCAGAGCATGTGTGTGTGTGTGTGTGTGTGTCTCTGT	32137
QY	316	GTCTCTCACTCTT	375
DB	32136	GTCTCTCTCTT	32077
QY	376	TCATCTATCATTTACATAGTAGTATTAATAATAATTAGAGAGATACAGAAA	430
DB	32076	TCACCTGCATAGTATTAAACATTTTCAAAGTTACTTGCACCAATCTAGAAAGATA	32022

	Query Match	50.9%	Score 339.2	DB 2	Length 171627
	Best Local Similarity 87.1%	Pred. No. 6.3e-78			
	Matches 399	Conservative 0	Mismatches 48	Indels 11	Gaps 2
Qy	16	GAGGTCCATAACCCCTGCCCTCATCCAGATC---	TGTGCAGATGAAGAGAGGGAGGGA	71	
Db	101200	GGGACCATTAACCCCTGCCCTCATCCAGATCTGTGTGCAGATGAAGAGAGGGAGGGA	101141		
Qy	72	GAGGGAAGAGAGAGATGCTTTGGGGTGATTTGGCCAGAGGCCACCGCTGGATCCCA	131		
Db	101140	GAGGGAAGAGAGAGATGCTTTGGGGTGATTTGGCCAGAGGCCACCGCTGGATCCCA	101081		
Qy	132	TGAAGAAATCTGGGTGAGAGGGTCTTAAAGTCATAAACTGAGATCCAGTTCGCAGGTGGC	191		
Db	101080	TGAAGAAATCTGGGTGAGAGGGTCTTAAAGTCTATAAACTGAGATCCAGTTCGCAGGTGGC	101021		
Qy	192	TGCATAGTTGCCAACACAGTGAATGTGTCACTTTTGATCTTTCATCAGAATCTCAGCGTG	251		
Db	101020	TGCATAGTTGCCAACACAGTGAATGTGTCACTTTTGATCTTTCATCAGAATCTCAGCGTG	100961		
Qy	252	GTGCCACCTGGGCCAAATACACTGCAGAGCATCTGTCTGTCTGTCTGTCTGTCTCT	311		
Db	100960	GTGCCACCTGGGCCAAATACACTGCAGAGCATCTGTCTGTCTGTCTGTCTGTCTCT	100901		
Qy	312	CTGTCTCTCAC	371		
Db	100900	CTGTCCC	100841		

Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpachy, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshuwa, L., Louised, H., Lozada, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmood, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mahoney, S., McLeod, M., McNeill, T., Meenen, E., Milosavljevic, A., Mirer, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwakaeme, O., Okwuonu, G., Olariunpugoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sigson, I., Sitter, C. D., Smajs, D., Sneed, A., Sodergren, E., Song, X., Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Tingey, A., Trejes, Z., Usmani, K., Valas, R., Vera, V., Villanana, D., Waldron, L., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

Direct Submission
Unpublished
2 (bases 1 to 139973)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (13-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GOSY
Center clone name: CH210-329K10
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 65209 bases at least Q40
Consensus quality: 71948 bases at least Q30
Consensus quality: 77254 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently consists of 66 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence. * as soon as it is available and the accession number will be preserved.
* 1 1002: contig of 1002 bp in length
* 1003 1102: gap of unknown length
* 1103 2267: contig of 1165 bp in length
* 2268 2367: gap of unknown length
* 3779: contig of 1412 bp in length
* 3780 3879: gap of unknown length
* 3880 5068: contig of 1189 bp in length
* 5069 5168: gap of unknown length
* 5169 6321: contig of 1153 bp in length
* 6321: contig of 1085 bp in length
* 7506: gap of unknown length
* 9197: contig of 1591 bp in length
* 9297: gap of unknown length
* 10744 10843: gap of unknown length
* 11871: contig of 1028 bp in length
* 11971: gap of unknown length
* 13021: contig of 1050 bp in length
* 13121: gap of unknown length
* 13122 14340: contig of 1219 bp in length
* 14341 14440: gap of unknown length
* 14441 15454: contig of 1014 bp in length
* 15455 15854: gap of unknown length
* 15855 16945: gap of unknown length
* 16946 18080: contig of 1135 bp in length
* 18081 18180: gap of unknown length
* 18181 19692: contig of 1512 bp in length
* 19693 19792: gap of unknown length
* 19793 21510: contig of 1718 bp in length
* 21511 21610: gap of unknown length
* 21611 22627: contig of 1017 bp in length
* 22628 23448: contig of 1121 bp in length
* 23449 23948: gap of unknown length
* 23949 25018: contig of 1070 bp in length
* 25019 26230: contig of 1112 bp in length
* 26231 26330: gap of unknown length
* 26331 27733: contig of 1403 bp in length
* 27734 27833: gap of unknown length
* 27834 29269: contig of 1436 bp in length
* 29270 29369: gap of unknown length
* 29370 30592: contig of 1223 bp in length
* 30593 31885: contig of 1193 bp in length
* 31886 33219: contig of 1234 bp in length
* 33220 35164: contig of 1845 bp in length
* 35165 36759: contig of 1495 bp in length
* 36760 36859: gap of unknown length
* 36860 38180: contig of 1321 bp in length
* 38181 38280: gap of unknown length
* 38281 39503: contig of 1223 bp in length
* 39504 39603: gap of unknown length
* 39604 41380: contig of 1777 bp in length
* 41381 41480: gap of unknown length
* 41481 42886: contig of 1406 bp in length
* 42887 42986: gap of unknown length
* 42987 44735: contig of 1749 bp in length
* 44736 44835: gap of unknown length
* 44836 46342: contig of 1507 bp in length
* 46343 46443: contig of 2022 bp in length
* 46444 48564: gap of unknown length
* 48565 49989: contig of 1425 bp in length
* 49990 50089: gap of unknown length
* 50090 51487: contig of 1398 bp in length
* 51488 51587: gap of unknown length
* 51588 53757: contig of 2170 bp in length
* 53758 53857: gap of unknown length
* 53858 55205: contig of 1348 bp in length
* 55206 55305: gap of unknown length
* 55306 57182: contig of 1877 bp in length
* 57183 57282: gap of unknown length
* 57283 58631: contig of 1348 bp in length
* 58631 60121: contig of 1391 bp in length
* 60122 60221: gap of unknown length

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

[illegible]

```

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: zC218021
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 209402 bases at least Q40
Consensus quality: 209750 bases at least Q30
Consensus quality: 210020 bases at least Q20
Insert size: 21118; sum-of-contigs
Insert size: 208675; 5.8% error; agarose-fp
Quality coverage: 8.52x in Q20 bases;
Quality coverage: 8.81x in Q20 bases; agarose-fp

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

2651:	contig of 2651 bp in length
2751:	gap of 100 bp
12194:	contig of 118743 bp in length
12194:	contig of 100 bp
121594:	gap of 100 bp
128045:	contig of 6451 bp in length
128145:	gap of 100 bp
138046:	128145: gap of 100 bp
138146:	130387: contig of 2242 bp in length
138146:	130487: gap of 100 bp
130388:	130487: contig of 8101 bp in length
130488:	211588: contig of 8101 bp in length

```
FEATURES
source
1. 211588
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-218Q21"
/clone_lib="CHORI-211"
```

misc feature

```
misc_reduce
T: 2001
/note="assembly_fragment:00052
clone end:SP6
```

```
misc feature      2752, .121494
vector_side:left" clone end:350
```

```
misc_leaure
2/32. .121494
/note="assembly_
fragment chain:1"
```

```
misc_feature
121595. :128045
fragment_chain:1
/note="assembly fragment:01233
```

```
misc feature
128146. :130387
fragment_chain:1"
/notes/assembly/fragment.v12333
```

```
misc_record
140140: 150007
/note="assembly_fragment:00992
fragment chain:1"
```

```
misc_feature
130488. 211588
fragment_chain:1
/note="assembly fragment:01224
```

```

/note-assembly-in
fragment_chain:1
clone_end:T7
vector_side:right"

```

ORIGIN

```
Query Match      13.0%; Score 85.6; DB 2; Length 211588;
Best Local Similarity 73.8%; Pred. No. 4.3e-12;
Matches 110; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
```

[illegible]

RESULT 7	AL732512	46643 bp	DNA	linear	ROD 30-JUL-2002
LOCUS	AL732512/c				
DEFINITION	Mouse DNA sequence from clone Rp23-21G12 on chromosome 4, complete sequence.				
ACCESSION	AL732512				
VERSION	AL732512.7				
KEYWORDS	HTG.				

KEYWORDS
 Mus. musculus (house mouse)
SOURCE
ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 46643)
 1 (bases 1 to 46643)
 Philimore, B.
 Direct Submission
 Submitted (30-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humuquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Jul 31, 2002 this sequence version replaced at: 22003179.
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

COMMENT

 On Jul 31, 2002 this sequence version
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: hunquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

Only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-21G12 is

http://www.sanger.ac.uk/Projects/Chori/REGANS/working/REGANS.html
from the RPCI-23 Mouse PAC Library
constructed by the group of Pieter de Jong.
For further details see <http://www.chori.org/bacpac/home.htm>
VECTOR: pBACe3.6.

```

FEATURES
  Location/Qualifiers
    source
      1..46643
      /organism="Mus musculus"
      /mol_type="genomic DNA"
      /db_xref="txicon:10090"
      /chromosome="4"
      /clone="RP23-21G12"
      /clone_lib="pCI-29"

```

ORIGIN

```
Query Match      12.9%; Score 86.2; DB 10; Length 46643;
Best Local Similarity 59.2%; Pred. No. 5e-12;
Matches 183; Conservative 0; Mismatches 123; Indels 3; Gaps 2;
```

QY	393	AGTACTATAATAATAATATTAGAGATACACAGAAAATATATAGAGAGATAACAGTG	452
Db	14935	GAAGTAAAAATACAGTATTTCTATTAATAGTGACACATAAATAGATTATAGTCTCTG	14994
QY	453	TTCTCTATAAAAAAACAAGC	474
Db	14995	TACCATTAAAAAAGTCCAC	15016
RESULT	9		
AC115631			
LOCUS	AC115631	179355 bp	DNA linear ROD 05-NOV-2003
DEFINITION	Mus musculus BAC clone RP24-128K6 from 18, complete sequence.		
ACCESSION	AC115631		
VERSION	AC115631.4	GI:22539286	
KEYWORDS	HTG.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. Mus.		
AUTHORS	Goyea, E. and Haglund, K.		
TITLE	The sequence of Mus musculus BAC clone RP24-128K6		
JOURNAL	Unpublished (2001)		
REFERENCE	2 (bases 1 to 179355)		
AUTHORS	Wilson, R.		
TITLE	Sequencing of Mus musculus		
JOURNAL	Unpublished (2001)		
REFERENCE	3 (bases 1 to 179355)		
AUTHORS	McPherson, J.D. and Waterston, R.H.		
TITLE	Direct Submission		
JOURNAL	Submitted (21-MAR-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA		
REFERENCE	4 (bases 1 to 179355)		
AUTHORS	McPherson, J.D. and Waterston, R.H.		
TITLE	Direct Submission		
JOURNAL	Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA		
REFERENCE	5 (bases 1 to 179355)		
AUTHORS	McPherson, J.D. and Waterston, R.H.		
TITLE	Direct Submission		
JOURNAL	Submitted (29-AUG-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA		
REFERENCE	6 (bases 1 to 179355)		
AUTHORS	Wilson, R.		
TITLE	Direct Submission		
JOURNAL	Submitted (05-NOV-2003) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA		
COMMENT	On Aug 29, 2002 this sequence version replaced gi:22475842.		
	----- Genome Center		
	Center: Washington University Genome Sequencing Center		
	Center code: WUGSC		
	Web site: http://genome.wustl.edu		
	Contact: submissions@watson.wustl.edu		
	----- Summary Statistics		
	Center project name: M BB0128K06		

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION:

SOURCE INFORMATION:
The RPCI-24 BAC Library has been constructed by Pieter de Jong and coworkers (<http://www.chori.org>) from male C57BL/6J mouse spleen and/or brain genomic DNA. The clone and detailed information can be obtained from Pieter de Jong and coworkers at <http://www.chori.org>

NEIGHBORING SEQUENCE INFORMATION:

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone.

FEATURES
source

```
Location/Qualifiers
1. .179355
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="18"
/map="18"
/clone=RP24-128K6"
/clone.lib="RPC1-24"
```

[illegible]

repeat_region	/rpt_family="MaLR"	11565..11785
repeat_region	/rpt_family="B2"	11843..11979
repeat_region	/rpt_family="Alu"	12150..12262
repeat_region	/rpt_family="B4"	12614..12759
repeat_region	/rpt_family="Alu"	12804..12950
repeat_region	/rpt_family="Alu"	13140..13154
repeat_region	/rpt_family="B4"	13155..13288
repeat_region	/rpt_family="Alu"	13289..13305
repeat_region	/rpt_family="B4"	13303..13404
repeat_region	/rpt_family="Alu"	13463..13509
repeat_region	/rpt_family="SRV1"	13551..13777
repeat_region	/rpt_family="MaLR"	13778..13928
repeat_region	/rpt_family="Alu"	13925..14125
repeat_region	/rpt_family="MaLR"	15045..15192
repeat_region	/rpt_family="Alu"	15200..15378
repeat_region	/rpt_family="MaLR"	15396..15505
repeat_region	/rpt_family="B2"	15538..15699
repeat_region	/rpt_family="B2"	15725..15817
repeat_region	/product="RNA-Ser"	15871..15962
repeat_region	/rpt_family="MaLR"	15963..16130
repeat_region	/rpt_family="L1"	16315..16419
repeat_region	/rpt_family="Alu"	16357..16438
repeat_region	/rpt_family="B4"	16587..16724
repeat_region	/rpt_family="Alu"	16775..16877
repeat_region	/rpt_family="Alu"	16882..16967
repeat_region	/rpt_family="B4"	17039..17130
repeat_region	/rpt_family="ID"	17194..17506
repeat_region	/rpt_family="MaLR"	17508..17633
repeat_region	/rpt_family="Alu"	17650..17872
repeat_region	/rpt_family="B2"	18316..18537
repeat_region	/rpt_family="B2"	18590..18697

Query Match	12.8%	Score 85.2:	DB 10:	Length 179355;
-------------	-------	-------------	--------	----------------

Query Match 12.8%; score 85.2; DB 10;
Best local similarity 70.4%; pred. No. 9.9e-12;

Best Local Similarity	70.4%	Pred. No. 9.9e-12;	
Matches	114	Conservative	0
Mismatches	114	Mismatches	48
		Indels	0
		Gaps	0

[illegible]

[illegible]

RESULT 10	213715 bp	DNA	linear	HTG 20-OCT-2003
LOCUS	213715 bp	DNA	linear	HTG 20-OCT-2003
DEFINITION	Danio rerio clone DKEY-1954, WORKING DRAFT SEQUENCE, 9 unordered pieces.			
ACCESSION	EX649528			
VERSION	EX649528.2			
KEYWORDS	GI:37776948			
SOURCE	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.			
ORGANISM	Danio rerio (zebrafish)			
	Danio rerio			
	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi;			

REFERENCE	Cyprianinormes; Cypripinidae; Danio.
AUTHORS	1 (bases 1 to 213715)
TITLE	McLaren, S.
JOURNAL	Direct Submission
COMMENT	Submitted (18-OCT-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk On Oct-20, 2003 this sequence version replaced g1:35210297.

```

Center: wellcome illust sanger institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
----- Project Information -----
Center project name: zk19F4
----- Summary Statistics -----
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 209130 bases at least Q40
Consensus quality: 210003 bases at least Q30
Consensus quality: 210994 bases at least Q20
Insert size: 212915; sum-of-contigs
Insert size: 205869; 2.4% error; agarose-fp
Quality coverage: 10.08x in Q20 bases; sum-of-contigs
Quality coverage: 10.55x in Q20 bases; agarose-fp

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

*	5568	5667: gap of 100 bp
*	5688	3173: contig of 26086 bp in length
*	31754	31853: gap of 100 bp
*	31854	53919: contig of 22066 bp in length
*	53920	54019: gap of 100 bp
*	54020	70286: contig of 16367 bp in length
*	70287	70386: gap of 100 bp
*	70387	166490: contig of 86104 bp in length
*	156491	156590: gap of 100 bp
*	156591	169999: contig of 13409 bp in length
*	170000	170099: gap of 100 bp
*	170100	179893: contig of 9784 bp in length
*	179884	179983: gap of 100 bp
*	179984	187057: contig of 7074 bp in length
*	187058	187157: gap of 100 bp


```

RESULT 12
BX530074/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

```

BX530074      215778 bp    DNA    linear    HTG 29-SEP-2003
Danio rerio clone DKEY-266K15, WORKING DRAFT SEQUENCE, 7 unordered
pieces.
BX530074      GI:37051046
HTG: HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
Danio rerio (zebrafish)
Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 215778)
Melaren.S.
Direct Submission
Submitted (28-SEP-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfsh-help@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Sep 29, 2003 this sequence version replaced gi:32134827.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfsh-help@sanger.ac.uk
----- Project Information
Center project name: ZK66K15
----- Summary Statistics
Assembly program: GMAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 213339 bases at least Q30
Consensus quality: 214215 bases at least Q40
Consensus quality: 214727 bases at least Q20
Insert size: 215178; sum-of-contigs
Insert size: 198791; 5.1% error; agarose-ff
Quality coverage: 7.70x in Q20 bases; sum-of-contigs Quality
coverage: 8.44x in Q20 bases; agarose-ff
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
*
* 1 4918: contig of 4918 bp in length
* 4919 5018: gap of 100 bp
* 5019 71913: contig of 66895 bp in length
* 71914 72013: gap of 100 bp
* 72014 135787: contig of 53774 bp in length
* 135788 135887: gap of 100 bp
* 135888 189725: contig of 63838 bp in length
* 189726 189825: gap of 100 bp
* 189826 191911: contig of 2086 bp in length
* 191912 192011: gap of 100 bp
* 192011 206777: contig of 14766 bp in length
* 206778 206877: gap of 100 bp
* 206878 215778: contig of 8901 bp in length.
*
Location/Qualifiers
1..215778
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEY-266K15"
/clone_lib="DanioKey"
misc_feature 1..4918
/note="assembly fragment:01270
fragment_chain:1"
5019..71913
/note="assembly fragment:02287"

```

```

----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly: Phrap; version 0.990319
Consensus quality: 13243 bases at least Q40
Consensus quality: 13392 bases at least Q30
Consensus quality: 13354 bases at least Q20
Insert size: 24000; agarose-fp
Quality coverage: 16.32 in Q20 bases; agarose-fp
Quality coverage: 15.35 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1. 5756: contig of 5756 bp in length
* 5757 5856: gap of unknown length
* 12184 12184: contig of 6328 bp in length
* 12285 12284: gap of unknown length
* 25984 25984: contig of 13700 bp in length
* 26084 26084: gap of unknown length
* 45068 45068: contig of 18984 bp in length
* 45169 45169: gap of unknown length
* 79414 79414: contig of 34246 bp in length
* 79515 79515: gap of unknown length
* 79515 135904: contig of 56390 bp in length.
* Location/Qualifiers
* 1..135904
* /organism="Mus musculus"
* /mol_type="genomic DNA"
* /db_xref="taxon:10090"
* /chromosome="UNK"
* /clone="RP24-152G20"
* 1..5756
* /note="assembly_name:Contig18"
* 5857..12184
* /note="assembly_name:Contig19"
* 12285..25984
* /note="assembly_name:Contig20"
* 26085..45068
* /note="assembly_name:Contig21"
* 45169..79414
* /note="assembly_name:Contig22"
* 79515..135904
* /note="assembly_name:Contig23"
-----
ORIGIN
Query Match 12.7%; Score 84.8; DB 2: Length 135904;
Best Local Similarity 60.3%; Pred No. 1.2e-11;
Matches 140; Conservative 0; Mismatches 92; Indels 0; Gaps 0;
QY 236 CAGAAATCTCAGCCCTGGTGGCCACCTGCCAAATACACTGCAGAGCAATGCTGTCTCT 295
Db 83350 CAGAACAGGATCCCTGTACTGGATCTTGAAAACTCATTTGTGTGTGTGTGTGTGT 83291
QY 296 GTCTGTCTGTCTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 355
Db 83290 GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 83231
QY 356 GTGTGTGTCTCCTCACTCTTTTCATCCTATCATATACATAGTAGTATAATAATAATTAG 415
Db 83230 GTGTGTGTATATATATATATTTATCTATAGCCGTAAAAAGTAAAAAAGGTTAAAAATGC 83171
QY 416 AGAGATACACAGAAAAATATAGAGACATACAGTGTCTCTATAAAAAA 467

```

Db	83170	ACAGCTTTAAATTCAAATTTTCAAAAGATATCTATAACATCAGAACTAAA	83119
RESULT 14			
AC130218		213673 bp DNA linear	ROD 19-NOV-2003
LOCUS			
DEFINITION		Mus musculus chromosome 18 clone RP23-121N18, complete sequence.	
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			
ORIGIN			
Query Match			
Best Local Similarity			
Matches 140; Conservative			
Qy	236	CAGAAATCTCAGCTGTGTGCGCCACCTGGCCAAATACACTGCAGAGCATGCTGTGTGCT	295
Db	56329	CAGAACGAGTCCCTGACTGATCTTGAAAACTCATTTGTGTGTGTGTGTGTGTGTGT	56388
Qy	296	GTCTGTGTGTGTCTCTGT	355
Db	56389	GT	56448
Qy	356	GTGTGTGTCTCCCTCACCTCTTCATCCCTATCATATAGTAGTATATAATAATATTAG	415
Db	56449	GTGTGTGTAT	56508
Qy	416	AGAGATACACAGAAAATATATAGAGAATAACAGTGTCTCTATATAAAAAA	467
Db	56509	ACAGCTTTAAATTCAAATTTTCAAAAGATATCTATAACATCAGAACTAAA	56560
RESULT 15			
AC139385/c			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			
ORIGIN			
Query Match			
Best Local Similarity			
Matches 140; Conservative			
Qy	236	CAGAAATCTCAGCTGTGTGCGCCACCTGGCCAAATACACTGCAGAGCATGCTGTGTGCT	295
Db	56329	CAGAACGAGTCCCTGACTGATCTTGAAAACTCATTTGTGTGTGTGTGTGTGTGTGT	56388
Qy	296	GTCTGTGTGTGTCTCTGT	355
Db	56389	GT	56448
Qy	356	GTGTGTGTCTCCCTCACCTCTTCATCCCTATCATATAGTAGTATATAATAATATTAG	415
Db	56449	GTGTGTGTAT	56508
Qy	416	AGAGATACACAGAAAATATATAGAGAATAACAGTGTCTCTATATAAAAAA	467
Db	56509	ACAGCTTTAAATTCAAATTTTCAAAAGATATCTATAACATCAGAACTAAA	56560

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 21, 2004, 19:26:07 ; Search time 651.769 Seconds
(without alignments)
10930.600 Million cell updates/sec

Title: US-10-078-090-48
Perfect score: 1677
Sequence: 1 gacttcggtggtccaaagg.....aagcgagacaggggggaga 1677

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373663 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : N Geneseq_29Jan04:*
- 1: Geneseqn1980s:*
 - 2: Geneseqn1990s:*
 - 3: Geneseqn2000s:*
 - 4: Geneseqn2001as:*
 - 5: Geneseqn2001bs:*
 - 6: Geneseqn2002s:*
 - 7: Geneseqn2003as:*
 - 8: Geneseqn2003bs:*
 - 9: Geneseqn2003cs:*
 - 10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1677	100.0	1677	6	ABV83605 Human bre
2	647.2	38.6	667	6	ABV83604 Human bre
3	554.8	33.1	3102	3	AACT5441 Human ORF
4	420.2	25.1	1718	7	ADA53214 Human cod
5	108.6	6.5	1338	4	ABL29731 Drosophil
6	108.6	6.5	6497	4	ABL29730 Drosophil
7	107.6	6.4	2822	7	ABX71178 Novel hum
8	76.8	4.6	3316	4	ABLI10742 Drosophil
9	76.6	4.6	1817	5	ABAI5256 Human ner
10	75.6	4.5	271990	9	ADD25213 Fertility
11	75.2	4.5	710	6	ABT09598 Phase-1 R
12	73.8	4.4	1274	4	AH7317 Human PDZ
13	73.6	4.4	2278	4	ABLI12518 Drosophil
14	73.4	4.4	667	9	ADD19798 Seabass p
15	73.4	4.4	67832	8	ADA02801 Mouse itk
16	73.4	4.4	67832	9	ADD72539 Mouse itk
17	73.4	4.4	67832	9	ADC85281 Mouse itk
18	73.2	4.4	42998	7	ADA14747 Human rib
19	73.2	4.4	42999	6	AB565032 Invertebr
20	73.2	4.4	42999	9	AD61411 Human rib
21	73.2	4.4	742	6	AB576748 Frog enbr
22	72.6	4.3	338	9	ADD19844 Oreochrom
23	72.4	4.3	596	9	ADD19952 Oreochrom

C	24	72.4	4.3	814	9	ADD20177	Add20177 Oreochrom
	25	72.4	4.3	41907	8	ADA37416	Ada37416 Human rDN
C	26	72.2	4.3	534	9	ADD20617	Add20617 Oreochrom
C	27	72.2	4.3	110000	7	ABX16390_5	Continuation (6 of
C	28	72	4.3	448	2	AAI68682	Aat68682 Penaeus v
C	29	72	4.3	605	9	ADD20225	Add20225 Oreochrom
	30	72	4.3	95599	9	ADC85298	Adc85298 Human Egr
	31	72	4.3	95600	8	ADA02819	Ada02819 Mouse Sos
	32	72	4.3	96600	9	ADB72557	Adb72557 Mouse Sos
C	33	71.8	4.3	299	9	ADD19778	Add19778 Gadus mor
	34	71.8	4.3	475	4	AAI12688	Aai12688 Probe #26
	35	71.8	4.3	475	4	ABA54378	Aba54378 Human foe
	36	71.8	4.3	475	4	AAI34039	Aai34039 Probe #27
	37	71.8	4.3	475	4	ABA43925	Aba43925 Human bre
	38	71.8	4.3	475	4	ABA24157	Aba24157 Probe #26
	39	71.8	4.3	475	4	AAK28111	Aak28111 Human bon
	40	71.8	4.3	475	4	AAK02673	Aak02673 Human bra
	41	71.8	4.3	475	4	ABS27717	Abs27717 Human liv
	42	71.8	4.3	475	5	AAI02597	Aai02597 Probe #25
	43	71.8	4.3	475	6	ABS02604	Abs02604 Human gen
C	44	71.8	4.3	527	6	ABQ92801	Abq92801 Triticum
	45	71.8	4.3	546	4	AAI19220	Aai19220 Probe #91

ALIGNMENTS

RESULT 1
ABV83605
ID ABV83605 standard; cDNA; 1677 BP.

XX ABV83605;
XX

DT 06-DRC-2002 (first entry)

XX Human breast specific gene SEQ ID NO 48.

XX Human; breast; cytostatic; cancer; transgenic; gene therapy; vaccine;
XX gene; ss.

XX Homo sapiens.

XX WO200266605-A2.

XX 29-AUG-2002.

XX 14-FEB-2002; 2002WO-US004284.

XX 15-FEB-2001; 2001US-0268999P.

XX (DIAD-) DIADEXUS INC.

XX Salceda S, Macina RA, Hu P, Recipon H, Karra K, Cafferkey R;
XX Sun Y, Liu C;

XX WPI; 2002-713345/77.

XX New isolated breast specific nucleic acid molecules and polypeptides,
XX useful for identifying, diagnosing, monitoring, staging, imaging and
XX treating breast cancer and non-cancerous disease states in breast tissue.
XX Claim 1; Page 171-172; 25app; English.

XX The invention relates to human breast specific nucleic acids (I)
XX comprising: (a) a sequence encoding any one of 95 protein sequences
XX (ABP66614-ABP66708); (b) any one of 115 polynucleotide sequences
XX (ABV83558-ABV83672); (c) a molecule that selectively hybridizes to (a) or
XX (b); (d) a molecule having at least 60% sequence identity to (a) or (b).
XX The breast specific nucleic acid molecules, polypeptides and antibodies
XX are useful for identifying, diagnosing, monitoring, staging, imaging and
XX treating breast cancer and non-cancerous disease states in breast tissue.
XX They are also useful for producing transgenic animals and cells and
XX producing engineered breast tissue for treatment and research. The

transgenic animals are useful as animal model systems used in elaborating the biological function of the polypeptides, studying conditions and/or disorders associated with aberrant expression and in screening for compounds effective in ameliorating the conditions. The polynucleotides are useful for gene therapy and in vaccines

CC	Query Match	100.08;	Score 1677;	DB 6;	Length 1677;	
CC	Best Local Similarity	100.08;	Pred. No. 0;			
CC	Matches 1677;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
XX	1	GAGTTGCGGCGTGCAGAGCCGACGAGGGCTTGGGCTTCAGCATCCGTCGGGGCTCGGAG	60			
XX	1	GAGTTGCGGCGTGCAGAGCCGACGAGGGCTTGGGCTTCAGCATCCGTCGGGGCTCGGAG	60			
XX	61	CACGGCTGGGCGATCTACGTCTCTCTGTGTGAAACAGGCTCTCTAGCTGAGAGGAGGA	120			
XX	61	CACGGCTGGGCGATCTACGTCTCTCTGTGTGAAACAGGCTCTCTAGCTGAGAGGAGGA	120			
XX	121	CTGGGGTGGGGGACGAGATCTGGGCTCAAGCAAAATCCCTGGCCCGGGTGAACCCAC	180			
XX	121	CTGGGGTGGGGGACGAGATCTGGGCTCAAGCAAAATCCCTGGCCCGGGTGAACCCAC	180			
XX	181	CGGAGGCGCTCAAGGCTCTGAAAGGCTCCAAAGAGCTGTGTGTCTGTGTACTCAGCA	240			
XX	181	CGGAGGCGCTCAAGGCTCTGAAAGGCTCCAAAGAGCTGTGTGTCTGTGTACTCAGCA	240			
XX	241	GGGCGATCCCTGGGGCTACGTACCAACACATCTACACCTGGGTGGACCCGCGAGGC	300			
XX	241	GGGCGATCCCTGGGGCTACGTACCAACACATCTACACCTGGGTGGACCCGCGAGGC	300			
XX	301	CGGAGCATCTCCCAACCTCGGGCTGCCCCAGGCCCAACGGTGTGCTGTGTGTGTGTGT	360			
XX	301	CGGAGCATCTCCCAACCTCGGGCTGCCCCAGGCCCAACGGTGTGCTGTGTGTGTGTGT	360			
XX	361	GAGGTGACCGGAGGACACCTGCACCTCTGCAAGAGGGGATGAGAAAGGTGAAC	420			
XX	361	GAGGTGACCGGAGGACACCTGCACCTCTGCAAGAGGGGATGAGAAAGGTGAAC	420			
XX	421	CTGGTGTGGGGGACGGCGGCTCCCTGGGCTCTACGATCCCTGGGGAGCTGAGTACGGC	480			
XX	421	CTGGTGTGGGGGACGGCGGCTCCCTGGGCTCTACGATCCCTGGGGAGCTGAGTACGGC	480			
XX	481	CTTGCCATTTACATCACTGGGGTGAGCCAGGCTCTGAGCAGAGGAGGGGCTCAAG	540			
XX	481	CTTGCCATTTACATCACTGGGGTGAGCCAGGCTCTGAGCAGAGGAGGGGCTCAAG	540			
XX	541	GTGCGGACCCAGATTTAGAAAGTGAATGGCGGAGCTTTCTCAACATCTCTACACGAGG	600			
XX	541	GTGCGGACCCAGATTTAGAAAGTGAATGGCGGAGCTTTCTCAACATCTCTACACGAGG	600			
XX	601	GCTGTGAGGCTGCTTAAGTCTCTCGGCACTCTCTGAGCAGTGAAGGAGCTCGGAGG	660			
XX	601	GCTGTGAGGCTGCTTAAGTCTCTCGGCACTCTCTGAGCAGTGAAGGAGCTCGGAGG	660			
XX	661	CTGCCCCATGCCCCGACCACTGTGAGCAGAGCAAGTGGATCGCCGATCAGG	720			
XX	661	CTGCCCCATGCCCCGACCACTGTGAGCAGAGCAAGTGGATCGCCGATCAGG	720			
XX	721	GAGACCATGGGAACTCGGAGGGTCTGGCACTCTGCTGCTCAATCTCCAGACCCCA	780			
XX	721	GAGACCATGGGAACTCGGAGGGTCTGGCACTCTGCTGCTCAATCTCCAGACCCCA	780			
XX	781	GGGCGATTTCTGAAAGCAGTATAGCTTCCCATCCCTCCACCGGCTGTCTCTCT	840			
XX	781	GGGCGATTTCTGAAAGCAGTATAGCTTCCCATCCCTCCACCGGCTGTCTCTCTCT	840			
XX	841	CTCAGCTGCACTCCCCACACAGGGCCCTCCATTGGCAGGACATGACCTGGGACATCC	900			
XX	841	CTCAGCTGCACTCCCCACACAGGGCCCTCCATTGGCAGGACATGACCTGGGACATCC	900			
XX	901	CTCTCTCTCTTGGGCTCAGTTTCCCATGGAAAGCTGAAATACACCATCCAACTGTCTC	960			

Db	901	CTCTCTCTCTTGGGCTCAGTTTCCCATGGAAAGCTGAAATACACCATCCAACTGTCTC	960
Qy	961	ATTCTTTATTTGTCGCCAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	1020
Db	961	ATTCTTTATTTGTCGCCAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	1020
Qy	1021	AAAGTGGGACCATTAACCTGCGCTCATCCAGATCTGTGCAGATGAAGAGAGGGGG	1080
Db	1021	AAAGTGGGACCATTAACCTGCGCTCATCCAGATCTGTGCAGATGAAGAGAGGGGG	1080
Qy	1081	AGAGGGAAGAGAGAGATGCTTTGGGGTGTATTTGGCCAGAGGCCACAGGCTGGATCCC	1140
Db	1081	AGAGGGAAGAGAGAGATGCTTTGGGGTGTATTTGGCCAGAGGCCACAGGCTGGATCCC	1140
Qy	1141	ATGAAGAAATCTGGGTGAGAGGGTCTTAAAGTCAATAAATGAGATCCAGTGGCAGGTG	1200
Db	1141	ATGAAGAAATCTGGGTGAGAGGGTCTTAAAGTCAATAAATGAGATCCAGTGGCAGGTG	1200
Qy	1201	CTGCATAGTTGCAACAGTGAATGTGCACCTTTTGTATCTTCAACAGAAATCTCAGGCT	1260
Db	1201	CTGCATAGTTGCAACAGTGAATGTGCACCTTTTGTATCTTCAACAGAAATCTCAGGCT	1260
Qy	1261	GGTGGCACCTGGCCAAATACACTGCAGAGCATGTCTGTCTGTCTGTCTGTCTGTCTC	1320
Db	1261	GGTGGCACCTGGCCAAATACACTGCAGAGCATGTCTGTCTGTCTGTCTGTCTGTCTC	1320
Qy	1321	TCGT	1380
Db	1321	TCGT	1380
Qy	1381	CTCTTTTATCTTATCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	1440
Db	1381	CTCTTTTATCTTATCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	1440
Qy	1441	ATATATAGAGAGATTAACAGTGTCTCTATAAAAAAAGAGTGCCTCTCTGCATAG	1500
Db	1441	ATATATAGAGAGATTAACAGTGTCTCTATAAAAAAAGAGTGCCTCTCTGCATAG	1500
Qy	1501	CTTTAACTCTCAGCAACTCTTCGAGAAAAGAGCAAAAACGGGAGAAAACAAAGAAACA	1560
Db	1501	CTTTAACTCTCAGCAACTCTTCGAGAAAAGAGCAAAAACGGGAGAAAACAAAGAAACA	1560
Qy	1561	ACGGGAGACAGACTAGAGAAACACAGACAGCGGACAAAACACACGTGAGGAGCAACA	1620
Db	1561	ACGGGAGACAGACTAGAGAAACACAGACAGCGGACAAAACACACGTGAGGAGCAACA	1620
Qy	1621	CCAGAGGGCGAACCACATTAACCCACACACGTGAAAAAGCGAGACCGAGGGGGAGA	1677
Db	1621	CCAGAGGGCGAACCACATTAACCCACACACGTGAAAAAGCGAGACCGAGGGGGAGA	1677

RESULT 2
ABV83604
ID
ABV83604 standard; cDNA; 667 BP.
XX
AC
ABV83604;
XX
DT
06-DEC-2002 (first entry)
XX
DE
Human breast specific gene SEQ ID NO 47.
XX
KW
Human; breast; cytostatic; cancer; transgenic; gene therapy; vaccine;
XX
KW
gene; ss.
XX
OS
Homo sapiens.
XX
PN
WO20026605-A2.
XX
PD
29-AUG-2002.
XX
PF
14-FEB-2002; 2002WO-US004284.
XX

PR 15-FEB-2001; 2001US-0268999P.
 XX (DIAD-) DIADEXUS INC.
 PA Salceda S, Macina RA, Hu P, Recipon H, Karra K, Cafferkey R;
 PI Sun Y, Liu C;
 XX WPI; 2002-713345/77.
 DR
 XX New isolated breast specific nucleic acid molecules and polypeptides,
 PT useful for identifying, diagnosing, monitoring, staging, imaging and
 PT treating breast cancer and non-cancerous disease states in breast tissue.
 XX
 PS Claim 1; Page 170-171; 254pp; English.
 XX
 CC The invention relates to human breast specific nucleic acids (I)
 CC comprising: (a) a sequence encoding any one of 95 protein sequences
 CC (ABP66614-ABP66708); (b) any one of 115 polynucleotide sequences
 CC (ABV83558-ABV83672); (c) a molecule that selectively hybridizes to (a) or
 CC (b); (d) a molecule having at least 60% sequence identity to (a) or (b).
 CC The breast specific nucleic acid molecules, polypeptides and antibodies
 CC are useful for identifying, diagnosing, monitoring, staging, imaging and
 CC treating breast cancer and non-cancerous disease states in breast tissue.
 CC They are also useful for producing transgenic animals and cells and
 CC producing engineered breast tissue for treatment and research. The
 CC transgenic animals are useful as animal model systems used in elaborating
 CC the biological function of the polypeptides, studying conditions and/or
 CC disorders associated with aberrant expression and in screening for
 CC compounds effective in ameliorating the conditions. The polynucleotides
 CC are useful for gene therapy and in vaccines
 XX
 SQ Sequence 667 BP; 201 A; 140 C; 178 G; 148 T; 0 U; 0 Other;
 Query Match 38.6%; Score 647.2; DB 6; Length 667;
 Best Local Similarity 99.5%; Pred. No. 1.8e-153;
 Matches 649; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 1026 GGGGACATACCCCTGCCCTCATCCAGATCTGTGCAGATGAAGAGAGGGAGGAGG 1085
 16 GAGGTCCATAACCCCTGCCCTCATCCAGATCTGTGCAGATGAAGAGAGGGAGGAGG 75
 1086 GAAAGAGAGAGATGCTTGGGGTGATTTGGCCAGAGGCCACAGGCTGGATCCCATGAA 1145
 76 GAAAGAGAGAGATGCTTGGGGTGATTTGGCCAGAGGCCACAGGCTGGATCCCATGAA 135
 1146 GAAATCTGGGTGAGAGGGTCTTAAAGTCATAAAGTCAGATCCAGTTGCCAGTGGCTGCA 1205
 136 GAAATCTGGGTGAGAGGGTCTTAAAGTCATAAAGTCAGATCCAGTTGCCAGTGGCTGCA 195
 1206 TAGTTCACACAGTGAATGTGTACCTTTTGATCTTCATCAGAAATCTCAGCTGGTGG 1265
 196 TAGTTCACACAGTGAATGTGTACCTTTTGATCTTCATCAGAAATCTCAGCTGGTGG 255
 1266 CCACCTGGCCAAATACACTGCAGAGCATGTCTGTCTGTCTGTCTGTCTGTCTCTCT 1325
 256 CCACCTGGCCAAATACACTGCAGAGCATGTCTGTCTGTCTGTCTGTCTGTCTCTCT 315
 1326 GTCTCTCTCT 1385
 316 GTCTCTCTCT 375
 1386 TCATCTTATCATATACATAGTAGTATATAATAATATTAGAGAGATACACAGAAATATA 1445
 376 TCATCTTATCATATACATAGTAGTATATAATAATATTAGAGAGATACACAGAAATATA 435
 1446 TAGAGAGATAACAGTGTCTTATAAAAAAAGAGTGGCCCTCTCTGCATAGTCTTCT 1505
 436 TAGAGAGATAACAGTGTCTTATAAAAAAAGAGTGGCCCTCTCTGCATAGTCTTCT 495
 1506 AACAACTCAGCACTCTCGCAGAAAGAGCACAACCGGAGAGAAACAGAAACAAACGGG 1565
 496 AACAACTCAGCACTCTCGCAGAAAGAGCACAACCGGAGAGAAACAGAAACAAACGGG 555

QY 1566 AGACAAGACTAGAGAAAACACAGGACAGCGGAGCAAAAACACCGTAGGGAGCAACACCAGA 1625
 Db 556 AGACAAGACTAGAGAAAACACAGGACAGCGGAGCAAAAACACCGTAGGGAGCAACACCAGA 615
 QY 1626 GGGGCGAAGCACATTACCCACACACACGTGAAAAGGAGGAGACACAGGGGGGAGA 1677
 Db 616 GGGGCGAAGCACATTACCCACACACACGTGAAAAGGAGGAGACACAGGGGGGAGA 667
 RESULT 3
 AAC75441
 ID AAC75441 standard; cDNA; 3102 BP.
 XX
 AC AAC75441;
 XX
 DT 08-FEB-2001 (first entry)
 XX Human ORF ORF996 polynucleotide sequence SEQ ID NO:1991.
 DE Human; open reading frame; ORF; detection; cytostatic; hepatotropic;
 XX vulnary; antipapillary; antipapillary; antipapillary; antipapillary;
 XX anticovulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antinflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.
 XX Homo sapiens.
 OS WO2000058473-A2.
 XX
 PN 05-OCT-2000.
 PD 31-MAR-2000; 2000WO-US008621.
 PF 31-MAR-1999; 99US-0127607P.
 PR 02-APR-1999; 99US-0127636P.
 PR 05-APR-1999; 99US-0127728P.
 PR 30-MAR-2000; 2000US-00540763.
 XX (CURA-) CURAGEN CORP.
 XX Shimkets RA, Leach M;
 PI WPI; 2000-602362/57.
 DR P-PSDB; AAB41232.
 XX Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease.
 XX
 PS Claim 5; Page 1517-1519; 5507pp; English.
 XX AAC7446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORF open reading frames 1 to 3161. The ORF
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antipapillary; antipapillary; antipapillary; antipapillary;
 CC anticovulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 CC immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 CC antirheumatic; antinflammatory; antithyroid; and antianemic. The
 CC dermatological; immunosuppressive; antinflammatory; antibacterial;
 CC antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The
 CC sequences can be used for determining the presence of or predisposition
 CC to, or preventing or treating pathological conditions associated with an
 CC ORF-associated disorder. The nucleic acids can be used to express ORF
 CC proteins in gene therapy vectors. The proteins and nucleic acids may be
 CC used to treat cancers, proliferative disorders, neurodegenerative

AC ABL29730;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 40663.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
XX
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
PI WPI; 2001-656860/75.
XX
DR New isolated nucleic acid detection reagent for detecting 1000 or more
XX
PT genes from Drosophila and for elucidating cell signaling and cell-cell
XX
PT interactions.
XX
PS Claim 1; SEQ ID NO 40663; 21pp + Sequence Listing; English.
XX
SQ The invention relates to an isolated nucleic acid detection reagent
XX
CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX
CC useful in developmental biology and in elucidating cell signalling and
XX
CC cell-cell interactions in higher eukaryotes for the development of
XX
CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
XX
CC sequences (AB101840-AB116175) and the encoded proteins (AB557737-
XX
CC AB572072). The sequence data for this patent did not form part of the
XX
CC printed specification, but was obtained in electronic format directly
XX
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 6497 BP; 1933 A; 1221 C; 1417 G; 1926 T; 0 U; 0 Other;
Query Match 6.5%; Score 108.6; DB 4; Length 6497;
Best Local Similarity 60.8%; Pred. No. 1.1e-16;
Matches 177; Conservative 0; Mismatches 114; Indels 0; Gaps 0;
QY 415 GTGAACCTGTGCTGGGGACGGCGGTCCCTGGGCTCAGCATCCGTGGGGAGCTGAG 474
DB 4982 GTGGAGCTGTGATCGAACCTGTGCTGGGCTGATGATCCGTGGGGCTGGAG 4923
QY 475 TACGGCTTGGCATTTACATCACTGGGCTGAGCCAGGCTCTGAAGCAGAGGAGCGGG 534
DB 4922 TACGGCTTGGGATCTTTGTCCCGGCTAGATAGGACAGCGTGGCAGATCGATCCGGA 4863
QY 535 CTCAGGTTGGGACCAAGATTCTAGAGTGAATGGGGAGCTTCTCAACATCCTACAC 594
DB 4862 CTGATGATGGGACGAGATCTCTGAGTCAATGGGCAATCTTCTGATGTGACGAC 4803
QY 595 GACGAGGCTGTGAGGCTGCTTAAGTCACTCCGACCTCATCTCCCTGACAGTGAAGACGTC 654
DB 4802 GACGAGGCTGTGAGGCTGCTTAAGTCACTCCGACCTCATCTCCCTGACAGTGAAGACGTC 4743
QY 655 GGGAGGCTGCCCGATGCCCGGACCACTGTGGACGAGACCAAGTGGATCGCC 705
DB 4742 GGCAAGGTGCCCGACTCTCTGACCTCCATCGAGATGGAGCCCTGGGACGCC 4692

RESULT 7
ABL29730/c
ID ABL29730 standard; cDNA; 2822 BP.

XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 40666.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
XX
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
PI WPI; 2001-656860/75.
XX
DR New isolated nucleic acid detection reagent for detecting 1000 or more
XX
PT genes from Drosophila and for elucidating cell signaling and cell-cell
XX
PT interactions.
XX
PS Claim 1; SEQ ID NO 40666; 21pp + Sequence Listing; English.
XX
SQ The invention relates to an isolated nucleic acid detection reagent
XX
CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX
CC useful in developmental biology and in elucidating cell signalling and
XX
CC cell-cell interactions in higher eukaryotes for the development of
XX
CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
XX
CC sequences (AB101840-AB116175) and the encoded proteins (AB557737-
XX
CC AB572072). The sequence data for this patent did not form part of the
XX
CC printed specification, but was obtained in electronic format directly
XX
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1338 BP; 303 A; 400 C; 384 G; 251 T; 0 U; 0 Other;
Query Match 6.5%; Score 108.6; DB 4; Length 1338;
Best Local Similarity 60.8%; Pred. No. 5.4e-17;
Matches 177; Conservative 0; Mismatches 114; Indels 0; Gaps 0;
QY 415 GTGAACCTGTGCTGGGGACGGCGGTCCCTGGGCTCAGCATCCGTGGGGAGCTGAG 474
DB 268 GTGGAGCTGTGATCGAACCTGTGCTGGGCTGATGATCCGTGGGGCTGGAG 327
QY 475 TACGGCTTGGCATTTACATCACTGGGCTGAGCCAGGCTCTGAAGCAGAGGAGCGGG 534
DB 328 TACGGCTTGGGATCTTTGTCCCGGCTAGATAGGACAGCTGGCAGATCGATCCGGA 387
QY 535 CTCAGGTTGGGACCAAGATTCTAGAGTGAATGGGGAGCTTCTCAACATCCTACAC 594
DB 388 CTGATGATGGGACGAGATCTCTGAGTCAATGGGCAATCTTCTCGATGTGACGAC 447
QY 595 GACGAGGCTGTGAGGCTGCTTAAGTCACTCCGACCTCATCTGACAGTGAAGACGTC 654
DB 448 GACGAGGCTGTGAGGCTGCTTAAGTCACTCCGACCTCATCTGACAGTGAAGACGTC 507
QY 655 GGGAGGCTGCCCGATGCCCGGACCACTGTGGACGAGACCAAGTGGATCGCC 705
DB 508 GGCAAGGTGCCCGACTCTCTGACCTCCATCGAGATGGAGCCCTGGGACGCC 558

RESULT 6
ABL29730/c
ID ABL29730 standard; DNA; 6497 BP.

XX AC ABX71178;
XX XX 05-MAR-2003 (first entry)
XX DE Novel human cDNA sequence #403.
XX KW Human; gene; ss; nervous system disorder; peripheral neuropathy;
KW Huntington's disease; amyotrophic lateral sclerosis; haemophilia;
KW neurodegenerative disease; Parkinson's disease; Alzheimer's disease;
KW autoimmune disease; systemic lupus erythematosus; rheumatoid arthritis;
KW insulin-dependent diabetes mellitus; anaemia; thrombocytopaenia; wound;
KW ulcer; burn; bone disorder; osteoporosis; osteoarthritis; stroke;
KW fibrosis; reperfusion injury; infection; allergic rhinitis; asthma;
KW coagulation disorder; cancer; tumour; inflammatory disease; septic shock;
KW Crohn's disease; anaphylaxis; proliferation; chemotactic;
KW differentiation; stem cell growth factor; haematopoiesis; chemokinetic;
KW haemostatic; antiinflammatory; expressed sequence tag; EST.
XX OS Homo sapiens.
XX XX WO200281731-A2.
XX XX 17-OCT-2002.
XX PF 29-JAN-2002; 2002WO-US001222.
XX XX 30-JAN-2001; 2001US-0074528.
XX XX (HYSB-) HYSEQ INC.
XX XX (GOOD/) GOODRICH R W.
XX PI Tang TY, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
XX PI Xue AJ, Yang Y, Wehrman T, Wang J, Wang D, Drmanac RT;
XX XX WPI; 2003-058563/05.
XX PT Novel polypeptide useful for treating neurodegenerative diseases, myeloid
XX FT or lymphoid cell disorders, bone disorders, mechanical and traumatic
XX PT disorders, coagulation disorders, and inflammatory diseases.
XX XX Claim 1; Page; 612pp; English.
XX CC This invention relates to the cDNA sequences encoding an isolated novel
XX CC human polypeptide. The protein encoded by the nucleic acid of the
XX CC invention is useful for treating central and peripheral nervous system
XX CC diseases (e.g. peripheral neuropathy, Huntington's disease, amyotrophic
XX CC lateral sclerosis); neurodegenerative diseases (e.g. Parkinson's disease,
XX CC Alzheimer's disease); autoimmune disease (e.g. systemic lupus
XX CC erythematosus, rheumatoid arthritis, insulin-dependent diabetes mellitus)
XX CC ; myeloid or lymphoid cell disorders (e.g. anaemia and thrombocytopaenia)
XX CC ; wounds, ulcers, burns; bone disorders (e.g. osteoporosis,
XX CC osteoarthritis); mechanical and traumatic disorders (e.g. stroke, head
XX CC trauma); lung or liver fibrosis; reperfusion injury in various tissues;
XX CC bacterial, viral or fungal infections; allergic conditions such as
XX CC allergic rhinitis, asthma; coagulation disorders (e.g. haemophilia);
XX CC cancer and tumours; and inflammatory diseases (e.g. septic shock, Crohn's
XX CC disease, anaphylaxis). The protein may be used to inhibit the growth,
XX CC infection or function of infectious agents such as bacteria, fungi,
XX CC viruses, or to effect bodily characteristics, biorhythms or circadian
XX CC cycles of rhythms. The protein may also have
XX CC proliferation/differentiation, stem cell growth factor, haematopoiesis
XX CC regulation, immune stimulating or suppressing, chemotactic/chemokinetic,
XX CC haemostatic and thrombolytic, receptor/ligand, and antiinflammatory
XX CC activities. The cDNA sequences of the invention are useful for expressing
XX CC recombinant protein for analysis. The present sequence represents a novel
XX CC human cDNA sequence of the invention, this sequence is an expressed
XX CC sequence tag (EST) and was identified using subtractive hybridisation
XX SQ Sequence 2822 BP; 557 A; 893 C; 922 G; 450 T; 0 U; 0 Other;

Query Match 6.4%; Score 107.6; DB 7; Length 2822;
Best Local Similarity 49.9%; Pred. No. 1.4e-16;

Matches 361; Conservative 0; Mismatches 324; Indels 39; Gaps 2;
QY 25 GAGGGTTGGGTTTACGATCCGTGGGGCTCGAGACCGCGTGGGCACTTACGTGTCT 84
DB 348 GGGAGGCTGGGCTTACGCTGGCGGGGCTCAGAGCATGGGCTTCTTCTGTCAGC 407
QY 85 CTGGTGGAAACAGGCTCTTAGTGTAGAAAGGAGTCTGGGTGGGGACGAGATTCG 144
DB 408 AAGTGGAGAGGAGAGTGTGAGAGCGGCTGGCTGTGGTGGGGACAGATCAGC 467
QY 145 CGGCTCAACGACAAATCCCTGGCCCGGGTGACCCAGCGGAGGCGGTCAAGGCTCTGAAG 204
DB 468 GAGGTGAATGGGCTGAGCCCTGGAGAGCACCACCATCGGTAGGCGCCGCTAAAGGTGCTGAC 527
QY 205 GGCTCCAAAGAGTGGTGTCTGTGTACTCAGCAGGGGCGCATCCCTGGGGCTACGTC 264
DB 528 AGCAGAGCGCGCTGCACATGATGGTTCCGGCGATGGCGGTGTGCCGGGCACTCAAGTTC 587
QY 265 ACCAACCAATCTACACCTGGGTGGACCCCGAGGGCGGAGCATCTCCCAACCTCGGGC 324
DB 588 TCCAAGGAGAGACACGCTGGGTGGATGTGTGAATCGGCGCTGTGTAGTGGAGAAAGTGC 647
QY 325 CTGCCCCCGCCACCGGTGGTGGCTGAGCAGCAGGAGGTGACCGGAGGACACCCCTG 384
DB 648 GGTTCACACCCCTCC-----GACACAGCTCAGAAGATGGTGTCCGGCGCATCTGTCAC 701
QY 385 CACCTCCTCGCAGGAGGGGATGAGAAAAAGGTGAACCTGTGTGGGGGACGGCCCGTCC 444
DB 702 CTATAC-----ACAACTCCGACGACTTCTGC 728
QY 445 CTGGGCTCAGATCCGTGGGGGAGTGTAGTACGGCTTGGCATTTACATCATCTGGGCTG 504
DB 729 CTGGGCTTCAACATCCGTGGGGCAAGAGATTTGGGCTGGGCTATGTGTCCAAAGTG 788
QY 505 GACCCAGGCTCTGAAGCAGAGGAGCGGGCTCAAGGTTGGGACACGAGATTCTAGAAAGTG 564
DB 789 GACCATGTGGGCTGGCCGAGAGGATGGCATCAAGGTGGGGGACCAAGTCTTGGCAGCC 848
QY 565 AATGGCGGAGCTTTCTCAACATCTTACAGCAGAGGCTGTCAAGGTCTTAAAGTCACT 624
DB 849 AACGGTGTCAAGTTTGACGACATCAGCCACAGCAGGCGCTGGAGGTGTGNAAGGGCCAA 908
QY 625 CGGCACCTCATCTCTGACAGTGAAGGAGCTCGGAGGCTGCCCATGCCGACCACTGTG 684
DB 909 ACGCAGCATCTGTGACCATCAAGGAGACCGCGGCTATCTGCCTACAAGGAGATGTT 968
QY 685 GACGAGACCAAGTGGATCGCCAGTTCGCCGATCAGGAGACCATGGCGAACTCGGCGAGG 744
DB 969 TCTGAGTACTGTGGCTGGACCGACTGAGCAACGGGGTGTGTCAGCAGCTGTCCCGGCC 1028
QY 745 TCTG 748
DB 1029 TCTG 1032
RESULT 8
ABL10742
ID ABL10742 standard; cDNA; 3316 BP.
XX ABL10742;
XX AC ABL10742;
XX XX 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 26708.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ss.
XX XX Drosophila melanogaster.
XX OS Drosophila melanogaster.
XX XX WO200171042-A2.
XX XX 27-SEP-2001.
PD

[illegible]

[illegible]

515	GGGAGGCTGGGCTTACGCGTGGCGGGGGCTCAGACGATGGCGCTGGGCAATCTTCGTGCAGC	574
Db		
85	CTGTGGGAACCAGGCTCTCTAGCTGAGAAGAAAGGACTCGGGGTCTGGGGACACAGATTCTG	144
QY		
575	AACTGGAGGAAGCAGCAGCTGCAGACGCGGCTGGCTGTGCGTGGGGACAAGATCAG	634
Db		
145	CGGTCAAACGACAAATCCCTGGCCCCGGGTGACCCACGGGAGGCCGCTCAAGGCTCTGAAG	204
QY		
635	GAGGTGAATGGGCTGAGCCTGGAGAGCACCACCATGGTAGCGCCGTAAGGTGCTGACC	694
Db		
205	GGCTCCAAAGAAGCTGGTGCTGTGTGTACTCAGCAGGGCGGCATCCCTGGGGGCTACAGTC	264
QY		
695	AGCAGACGCCCGCTGCACATGATGTTTCGGCGGCATGGCCGCTGTGCCGGGCATCAAGTTC	754
Db		
265	ACCAACCACTATTACCGTGG	285
QY		
755	TCCAAGGAGAAGACCGTGG	775
Db		

RESULT 13

ABL12518/c
ID ABL12518 standard: cDNA: 2278 BP.

XX
AC ABL12518:

DT 26-MAR-2000

Drosophila melanogaster ex

KW Drosophila; developmental biology; cell signalling; insecticide;

XX
XX

XX

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----

Figure 1

	6
	6
	6
	6
	4
	3
	6
	6
X	6
X	6

PR 11-JUL-2000; 2000US-00614150.

PA (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;
yy

DR WPI; 2001-656860/75.
DR D-DEBB. XBBCE941E

XX
Df New listed under act of 1907

PT genes from *Drosophila* and for interactions

XX
pg
Claim 1: SEQ ID NO

XX The invention relat

CC capable of detecting
CC useful in development

cell-cell interactions

CC discloses genomic D
CC sequences (ABL01840

CC
CC
CC

[illegible]

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Best Local Similarity

1

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 21, 2004, 19:28:36 ; Search time 6476.91 Seconds
(without alignments)
11222.365 Million cell updates/sec

Title: US-10-078-090-48
Perfect score: 1677
Sequence: 1 gatttgcgcgtgccaggc.....aagcgagaccagggggggaga 1677

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.mam.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	836.4	49.9	132292	9	AL138895	AL138895 Human DNA
C 2	829.4	49.5	171627	2	AF336382	AF336382 Homo sapi
C 3	742.4	44.3	3566	9	AB040959	AB040959 Homo sapi
C 4	610.2	36.4	3178	10	AK122523	AK122523 Mus muscu
C 5	585	34.9	2763	10	AY227205	AY227205 Rattus no
C 6	554.8	33.1	2902	9	HS0800985	AL110228 Homo sapi
C 7	420.2	25.1	1718	6	AX714098	AX714098 Sequence
C 8	420.2	25.1	1718	9	AK056190	AK056190 Homo sapi
C 9	180	10.7	237260	2	AC131430	AC131430 Rattus no
C 10	180	10.7	333531	2	AC127863	AC127863 Rattus no
C 11	172.6	10.3	214370	10	AL683828	AL683828 Mouse DNA
C 12	171.8	10.2	219809	2	AF336379	AF336379 Mus muscu
C 13	108.6	6.5	16325	2	AC015144	AC015144 Drosophil
C 14	108.6	6.5	63934	2	AC010699	AC010699 Drosophil
C 15	108.6	6.5	110000	3	AE003536_1	Continuation (2 of
C 16	108.6	6.5	173509	3	AC010031	AC010031 Drosophil
C 17	108.6	6.5	216041	2	AC112524	AC112524 Drosophil
C 18	107.6	6.4	1805	9	BC029054	BC029054 Homo sapi
C 19	107.6	6.4	2037	9	AK026862	AK026862 Homo sapi
C 20	96.2	5.7	179862	2	BX470066	BX470066 Danio rer
C 21	96.2	5.7	131463	2	EX510367	EX510367 Danio rer
C 22	96.2	5.7	219657	2	EX470198	EX470198 Danio rer
C 23	94.6	5.6	191463	2	EX510367	EX510367 Danio rer
C 24	89.6	5.3	2093	5	BC051776	BC051776 Danio rer
C 25	87.8	5.2	139973	2	AC135372	AC135372 Rattus no
C 26	86.6	5.2	211588	2	EX537348	EX537348 Danio rer
C 27	86.2	5.1	46643	10	AL732512	AL732512 Mouse DNA
C 28	86.2	5.1	214625	2	AC116582	AC116582 Mus muscu
C 29	85.2	5.1	168893	9	AF002987	AF002987 Homo sapi
C 30	85.2	5.1	179355	10	AC115631	AC115631 Mus muscu
C 31	85.2	5.1	213715	2	EX649528	EX649528 Danio rer
C 32	85	5.1	130176	2	EX005295	EX005295 Danio rer
C 33	85	5.1	203271	2	EX322566	EX322566 Danio rer
C 34	85	5.1	215778	2	EX530074	EX530074 Danio rer
C 35	84.6	5.0	117886	2	AC139385	AC139385 Rattus no
C 36	84.6	5.0	160718	2	EX664714	EX664714 Danio rer
C 37	84.6	5.0	202878	2	EX511101	EX511101 Danio rer
C 38	84.6	5.0	228962	2	AC131587	AC131587 Mus muscu
C 39	84.6	5.0	229574	2	AC123856	AC123856 Mus muscu
C 40	84	5.0	169095	5	EX088685	EX088685 Zebrafish
C 41	84	5.0	230877	5	EX248500	EX248500 Zebrafish
C 42	84	5.0	237343	2	EX569792	EX569792 Danio rer
C 43	83.8	5.0	163993	5	AL772362	AL772362 Zebrafish
C 44	83.6	5.0	135904	2	AC139157	AC139157 Mus muscu
C 45	83.6	5.0	142803	2	AC013771	AC013771 Homo sapi

ALIGNMENTS

RESULT 1
AL138895/c
LOCUS Human DNA sequence from clone RP11-9M16 on chromosome 9, complete
DEFINITION linear PRI 27-JUN-2001
ACCESSION AL138895
VERSION AL138895.16 GI:14575083
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 132292)
AUTHORS Skuce, C.
TITLE Direct Submission

Pred. No. is the number of results predicted by chance to have a

	repeat_region	30658.. .30770	/notes="L2 repeat: matches 2572. .2590 of consensus"
	repeat_region	31123.. .31357	/notes="L2 repeat: matches 2514. .2731 of consensus"
	repeat_region	32095.. .32170	/note="19 copies 4 mer acac 88% conserved"
	repeat_region	32365.. .32396	/note="16 copies 2 mer tc 84% conserved"
	repeat_region	32426.. .32485	/note="MIR repeat: matches 96. .152 of consensus"
	repeat_region	33041.. .33127	/note="MIR repeat: matches 20. .115 of consensus"
	repeat_region	33494.. .33764	/note="AluJ repeat: matches 4. .281 of consensus"
	repeat_region	34352.. .34417	/note="L2 repeat: matches 2641. .2709 of consensus"
	repeat_region	34554.. .34848	/note="AluX repeat: matches 1. .293 of consensus"
	repeat_region	34946.. .35142	/note="MIR repeat: matches 8. .207 of consensus"
	repeat_region	35336.. .35738	/note="AluS repeat: matches 1. .303 of consensus"
	repeat_region	35834.. .35965	/note="MIR repeat: matches 2. .147 of consensus"
	repeat_region	36769.. .36954	/note="MER53 repeat: matches 1. .188 of consensus"
	repeat_region	37658.. .37805	/note="MIR repeat: matches 40. .190 of consensus"
	repeat_region	37855.. .37962	/note="MIR repeat: matches 49. .140 of consensus"
	repeat_region	39025.. .39247	/note="MIR repeat: matches 14. .252 of consensus"
	repeat_region	39184.. .39263	/note="L2 repeat: matches 2671. .2750 of consensus"
	repeat_region	39516.. .39569	/note="MIR repeat: matches 98. .154 of consensus"
	repeat_region	40503.. .40600	/note="MIR repeat: matches 30. .126 of consensus"
	repeat_region	40730.. .40929	/note="MIR repeat: matches 46. .146 of consensus"
	repeat_region	40854.. .40893	
	 Query Match 49.9%; Score 836.4; DB 9; Length 132292; Best Local Similarity 95.5%; Pred.No. 2.7e-183; Matches 861; Conservative 0; Mismatched 41; Indels 0; Gaps 0;		
QY	539	AGGTGGGACACAGATTCTAGAAGTGATGGCGGAGCTTTCTCAATCCTCAACACG 598	
Dd	32923	AGTTGGGACCAGATTCTAGAAGTGATGGCGGAGCTTTCTCAATCCTCAACACG 32864	
QY	599	AGGCTCTCAGGTCGTTAAGTCACTTGGCACCTCATCTGCAGCTGAAGACCTCGGA 658	
Dd	32863	AGGCTCTCAGGTCGTTAAGTCACTTGGCACCTCATCTGCAGCTGAAGACCTCGGA 32804	
QY	659	GGTGCCCCATGCCGCACACTGTGGAGACCAAGTGGATCCCAAGTTCCTCGGATCA 718	
Dd	32803	GGTGCCCCGTGCCGCACCACTGTGGAGACCAAGTGGATCCCAAGTTCCTCGGATCA 32744	
QY	719	GGGAGACCATGGCGAACTCGCAGGCTCTGGCCACTCTGCTCGCTCCAATCTCCAGACC 778	
Dd	32743	GGGAGACCATGGCGAACTCGCAGGCTCTGGCCACTCTGCTCGCTCCAATCTCCAGACC 32684	
QY	779	CAGGCCATTTCTGAAGCCAGTGATAGTGCCTCCCATTCCCTCCACCGCCTGGCTCTC 838	
Dd	32683	CAGGCCATTTCTGAAGCCAGTGATAGTGCCTCCCATTCCCTCCACCGCCTGGCTCTC 32624	
QY	839	CTCTCAGCTTGCAGTCCCAACACAGGGCCCTCCATTGGCAGGACATGACCTGGGCAAT 898	
Dd	32623	CTCTCAGCTTGCAGTCCCAACACAGGGCCCTCCATTGGCAGGACATGACCTGGGCAAT 32564	
QY	899	CCCTCTCCCTCTTTGGCTCAGTTTTCCCATGGAAGACTGAAATACACCATCAACTGTC 958	
Dd	32563	CCCTCTCCCTCTTTGGCTCAGTTTTCCCATGGAAGACTGAAATACACCATCAACTGTC 32504	

QY	1435	CAGAAATATATACAGACGATACACAGTGTCTCTATAAAAAAAA	1479
DBD	100794	TTGCAACATCTAGAAGATACAGGTTTCTATAAAAAAAA	100750
RESULT 3			
LOCUS	AB040959		
DEFINITION	Homo sapiens mRNA for KIAA1526 protein, partial cds.		
ACCESSION	AB040959		
VERSION	AB040959.1	GI:7959318	
KEYWORDS	Homo sapiens (human)		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Artiodactyla; Homnidae; Homo.		
REFERENCE	1 (sttes), Kikuno, R., Ishikawa, K., Hirose, M. and Ohara, O.		
AUTHORS	Prediction of the coding sequences of unidentified human genes.		
TITLE	xvii. The complete sequences of 100 new cDNA clones from brain		
	which code for large proteins in vitro		
JOURNAL	DNA Res. 7 (2), 143-150 (2000)		
MEDLINE	20277482		
PUBMED	10819331		
REFERENCE	2 (bases 1 to 3566)		
AUTHORS	Chara, O., Nagase, T. and Kikuno, R.		
TITLE	Direct Submission		
JOURNAL	Submitted (04-APR-2000) Osamu Ohara, Kazusa DNA Research Institute,		
	Department of Human Gene Research; 1532-3, Yana, Kisarazu, Chiba		
	292-0812, Japan (E-mail: cdnainfo@kazusa.or.jp,		
	URL:http://www.kazusa.or.jp/huge, Tel:81-438-52-3913,		
	Fax:81-438-52-3914)		

```

FEATURES             Location/Qualifiers
     1..3566
     /organism="Homo sapiens"
     /mol_type="mRNA"
     /db_xref="taxon:9606"
     /clone="fj04743"
     /clone_lib="pBlueScriptII SK plus"

     1..3566
     /gene="KIAA1526"
     <1..2893
     /gene="KIAA1526"
     /note="Start codon is not identified. fj04743 cDNA clone
for KIAA1526 has a 196-bp insertion after the position
1335 of the sequence of KIAA1526."

     /codon_start=2
     /product="KIAA1526 protein"
     /protein_id="BAA96050.1"
     /db_xref="GI:7959319"
     /translation="GPRPPRQLPYCPSPNPGSSRRRLRAPGDPAVPVAPATAAPPTP
RALLRPGIVGEMNAPLDGLSVSSSTGSLGSAAGAGGAGLFLLSANVRQLHQAL
TALLSAEQREQTHCLNAYHARNVDFLTVLLDVPFKRLLPLMLRVIPRSDQL
LPDQVTEGLXPATTPYVROPAWGGDSAGPGEVFLSLRAKAHEGLGSGFSLGVSEH
GVGIVYSLVEKPIAEKSLGVGDQDILRVNOKSLARVTHAEAKVALKSKLVLGSEH
AGRIIPGVYNIHYTWDPDQGRSISPPSGLPQPHGGALRQOGDPSRSTHLLQGQDEK
KNLVLDGGRSLGTLRTGAEGVGLGIYITGVDPGSAEGSLKVGDDQLIENVNGSRFLN
ILHDVAURLKSSRHLLITKDVGRLPHARTITVDETKWIASRIETNANVNGSRFLN
TTTEGINKPQYFGPAGSVQTTSSLGNQTVLLEQVHLLNEQEHATWAYIYLDTEYRG
TTEGINKPQYFGLNMLTKAFLSEVNGTTSPODLRFDLHLVLRRETNESMKARQPPG
PGADGTYSWVSYSDTGSSTGSGHGTSTVSSARNTLDLETGEAVQGNINALPDSVDD
VSTDSOGLSFKELPRPPELPAQGNLPLQOGKRLGREDLPSSMSPGSCGTFVSAPQV
RPPAGTAPCTGTSAAQDLPSAPIYASVPANPSKRPDLDAHLALVNOHPIGPPRPVQ
SPFLHKSRAEATVAGGCLLPPSPGSHPDQGTNOHFVYVVEVHRDPSDSDYNEVRALP
QNTKATLSQLSDSQTCLLPPSPGSHPDQGTNOHFVYVVEVHRDPSDSDYNEVRALP
DGNAPKPEGLLEPTSLTVRVKKSAAATLGIAEGGANTRQPLPRIVTQKQSGSAHNCGLQ
KGVHVLVYNGTLTRGKHEARAARITAEAFKTKDRDYIDFLVTEFNVL"

```

QY 4 TTGCGGCGTGCACAGGCCCAAGGGCTTTCAGGCTTCAGGCTTCGCTGGGGGCTCGGAGCAC 63
 Db 596 TTGCGGCGTGCACAGGCCCAAGGGCTTTCAGGCTTCAGGCTTCGCTGGGGGCTCGGAGCAC 655
 QY 64 GCGGTGGGCACTACGTCTCTCTGTGTGAACACAGGCTCTCTAGCTGAGAGGAGGACTG 123
 Db 656 GCGGTGGGCACTACGTCTCTCTGTGTGAACACAGGCTCTCTAGCTGAGAGGAGGACTG 715
 QY 124 CGGTTCGGGAGCAGATTCTGGGGGTGACACAAATCCCTGGCCGGGTGACCCAGCG 183
 Db 716 CGGTTCGGGAGCAGATTCTGGGGGTGACACAAATCCCTGGCCGGGTGACCCAGCG 775
 QY 184 GAGGCGCTCAAGGCTCTCAAGGGTCCAAAGAGCTGTGTGTCTGTGTACTCAGCAGG 243
 Db 776 GAGGCGCTCAAGGCTCTCAAGGGTCCAAAGAGCTGTGTGTCTGTGTACTCAGCAGG 835
 QY 244 CCATCTCCCTGGGGGTACGTACCAACACATCTACACCTGTGGGTGACCCGAGGGCGC 303
 Db 836 CCATCTCCCTGGGGGTACGTACCAACACATCTACACCTGTGGGTGACCCGAGGGCGC 895
 QY 304 AGCATCTCCCAACCCCTCGGGCTGCCCCAGGCCCAAGGCTGTGTGTCTGTGGCAGCAGG 363
 Db 896 AGCATCTCCCAACCCCTCGGGCTGCCCCAGGCCCAAGGCTGTGTGTCTGTGGCAGCAGG 955
 QY 364 GGTGACCGGAGGAGCAGCTCTCTGCAAGGAGGGATGAGAAAGGTGAACCTG 423
 Db 956 GGTGACCGGAGGAGCAGCTCTCTGCAAGGAGGGATGAGAAAGGTGAACCTG 1015
 QY 424 GTGCTGGGGGAGCGCGGCTCGGGCTCACGATCCGTGGGGAGCTGAGTACGCGCTT 483
 Db 1016 GTGCTGGGGGAGCGCGGCTCGGGCTCACGATCCGTGGGGAGCTGAGTACGCGCTT 1075
 QY 484 GCATTTATATCATCTGGCGTGGACCCAGGCTCTGAAGCAGAGGAGCGGCTCAAGTT 543
 Db 1076 GCATTTATATCATCTGGCGTGGACCCAGGCTCTGAAGCAGAGGAGCGGCTCAAGTT 1135
 QY 544 GGGGACCAAGATTAGAAAGTGAATGGGCGGAGCTTTCTCAACATCTCTACACGAGGCT 603
 Db 1136 GGGGACCAAGATTAGAAAGTGAATGGGCGGAGCTTTCTCAACATCTCTACACGAGGCT 1195
 QY 604 GTGAGCTGTCTAGTCTATCTGGGCACTCTGCAAGTCAAGTGAAGAGCTGGGAGCTG 663
 Db 1196 GTGAGCTGTCTAGTCTATCTGGGCACTCTGCAAGTCAAGTGAAGAGCTGGGAGCTG 1255
 QY 664 CCCCATGCCCGCACCATCTGTGAGAGACCAAGTGGATCCGCAAGTCCCGGATCAGGAG 723
 Db 1256 CCCCATGCCCGCACCATCTGTGAGAGACCAAGTGGATCCGCAAGTCCCGGATCAGGAG 1315
 QY 724 ACCATGGCGAACTCGGAGGCTCT 747
 Db 1316 ACCATGGCGAACTCGGAGGCTTT 1339

RESULT 4
 AK122523
 LOCUS AK122523 3178 bp mRNA linear ROD 15-MAR-2003
 DEFINITION Mus musculus mRNA for mKIAA1526 protein.
 ACCESSION AK122523
 VERSION AK122523.1 GI:28972777
 KEYWORDS FLI_CDNA.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1
 AUTHORS Okazaki, N., Kikuno, R., Ohara, R., Inamoto, S., Aizawa, H., Yuasa, S.,
 Nakajima, D., Nagase, T., Ohara, O. and Koga, H.
 TITLE Prediction of the coding sequences of mouse homologues of KIAA
 gene: II. The complete nucleotide sequences of 400 mouse
 KIAA-homologous cDNAs identified by screening of terminal sequences
 of cDNA clones randomly sampled from size-fractionated libraries
 JOURNAL DNA Res. 10, 35-48 (2003)
 REFERENCE 2 (bases 1 to 3178)

AUTHORS Okazaki, N., Kikuno, R., Nagase, T., Ohara, O. and Koga, H.
 TITLE Direct Submission
 JOURNAL Submitted (07-FEB-2003) Hisashi Koga, Kazusa DNA Research
 Institute, Laboratory for Genome Informatics; 2-6-7
 Kazusa-kametari, Kisarazu, Chiba 232-0818, Japan
 (E-mail: mouse@kazusa.or.jp, Tel. 81-438-52-3919, Fax: 81-438-52-3918)
 COMMENT The CREATE program supported by Japan science and technology
 corporation; cDNA full insert sequencing; Kazusa DNA Research
 Institute; cDNA library construction; clone selection and 5' &
 3'-end one pass sequencing.
 FEATURES
 Location/Qualifiers
 1..3178
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="mh03417"
 /tissue_type="brain"
 /dev_stage="adult"
 /note="vector:modified pBC SK+"
 gene
 1..3178
 /gene="mKIAA1526"
 <1..2610
 /gene="mKIAA1526"
 /note="CDS is predicted by in silico analysis. Start codon
 is not identified."
 /evidence=not experimental
 /codon_start=1
 /product="mKIAA1526 protein"
 /protein_id="BAC65805.1"
 /db_xref="GI:28972778"
 /translation="ANVROLHQAALTALLSEPEREOTHCINAVHAPRVFDLVRVLRV
 LLSPVYKRLIPMLRIVIPRSDDLLFDQTAEGLYLPATIPYQPAWAAPDAGSGEV
 RLVRARAKAHEGLSGVSVISVEPGSLAEKRLVGGVDDGILLRNOKSL
 ARVTHABAKALGSKSLVSVYAGRIPGVYVTHIYVDPQGGSPSPSSLPQPH
 GSTLRQEDRRSTLLHSGDEKLVNLVIGDRLGSLTRGAEYGLGIYITGVDPG
 SEASSGLKVDQILVNGSFLNLIDAVKLLSRHLLITVDVGLPHARTVD
 QKWIASRIGESVANSAGFPDHTGEGTSKPGFYKPGAGSVTLSSLCNQTRALLDD
 QARHLTEQERATMYLAQYRGTTISVAMVMALPELLNTHAKTSLSEVSIISPO
 DLDFQHLVLERIESMKARQPPGQVGTVMVSYSDTSGTSGHSTSTVSSARNT
 LDLEGGETTQGSTNALPVSVDVDSPEDELPKIPPPPPPLAQCHDRILLGQPRKP
 GRDPLPSAAHSGIVFAPNRNRPPTGATPGSSADSDSPSPFIYASISHANPSS
 RLDPLHLALVNOHPFGPPVQSPHLLKSPFAETPGACILPPSPSPHPDAVGANQ
 HFVLVHRPDSPEVNEVRALPQTRTSLSDSGSLSDSDSDGADGETASTSGR
 QGTATKANKNGKEQPTREAGANKPFLLEPTSLVRVRKSAATLGIATGGANT
 RQPLRIVITIQRGSAHNCQLKVGHVILEVNGQTLRGRKEKARI IAEAFKTRER
 YIDPLVTEFNVML"
 ORIGIN
 Query Match 36.4%; Score 610.2; DB 10; Length 3178;
 Best Local Similarity 88.8%; Pred. No. 4.9e-131;
 Matches 660; Conservative 0; Mismatches 83; Indels 0; Gaps 0;
 QY 5 TCGCGCGTGCACAGGCCCAAGGGCTTTCAGGCTTCAGGCTTCGCTGGGGGCTCGGAGCAC 64
 Db 320 TCGCGCGTGCACAGGCCCAAGGGCTTTCAGGCTTCAGGCTTCGCTGGGGGCTCGGAGCAC 379
 QY 65 GCGTGGGCACTACGTCTCTCTGTGGAACACAGGCTCTCTAGCTGAGAGGAGGACTG 124
 Db 380 GCGTGGGCACTACGTCTCTCTGTGGAACACAGGCTCTCTAGCTGAGAGGAGGACTG 439
 QY 125 GCGTGGGCACTACGTCTCTCTGTGGAACACAGGCTCTCTAGCTGAGAGGAGGACTG 184
 Db 440 GCGTGGGCACTACGTCTCTCTGTGGAACACAGGCTCTCTAGCTGAGAGGAGGACTG 499
 QY 185 AGGCGGTCAAGGCTCTCAAGGGTCCAAAGAGCTGTGTGTCTGTGTACTCAGCAGG 244
 Db 500 AGGCGGTCAAGGCTCTCAAGGGTCCAAAGAGCTGTGTGTCTGTGTACTCAGCAGG 559
 QY 245 GCATCTCCCTGGGGGTACGTACCAACACATCTACACCTGGGTGGACCCGAGGGCGCA 304
 Db 560 GCATCTCCCTGGGGGTACGTACCAACACATCTACACCTGGGTGGACCCGAGGGCGCA 619
 QY 305 GCATCTCCCTGGGGGTACGTACCAACACATCTACACCTGGGTGGACCCGAGGGCGCA 364

Db 620 GCACATCCCTCCCTCCAGCCTGCCCCAGCCCATGGCAGCACCTTGAGACGGGTGAAG 679
QY 365 GTGACCGGAGGAGCACCTCGACCTCTGCAAGGAGGGGATGAGAAAAGGTGAACCTGG 424
Db 680 ATGACCGAAGGAGTACCTCCACCTCTCGAGGTGGAGATGAGAAAAGGTGAACCTGG 739
QY 425 TGCTGGGGAACGGCGGTCCTCGGCGCTCAACATCCGTGGGGAGCTGAGTACGGCTTG 484
Db 740 TGTGGGGGACGGCGGTCCTTGGCGCTCACGATCCGAGGTGGAGCAGAGTACGGCTTG 799
QY 485 GCATTATACATCACTGGCGTGGACCCAGGCTCTGAAGCAGAGGAGCAGCGGCTCAAGTTG 544
Db 800 GAATTATACATCACTGGTGGACCCAGGCTCTGAAGCAGAGGAGCAGCGGCTCAAGTTG 859
QY 545 GGGACCGAGTTCTAGAAGTGAATGGCGGAGCTTCTCAACATCTCAACAGCAGAGGCTG 604
Db 860 GAGACCGAGTTCTGAGAGTGAATGGCGGAGCTTCTCAACATCTCAACAGCAGAGGCTG 919
QY 605 TCAGGCTGCTTAAGTCACTCGGCACCTCATCTGACAGTGAAGCAGCTCGGGAGCTTGC 664
Db 920 TGAAGTGTCTCAAGTCACTCCCGCACCTCATCTGACGCTGAAGCAGCTCGGAGCTTGC 979
QY 665 CCATGCCCGACCACTGTGGACGAGACCAAGTGAATGCCAGTTCGGGATCAGGGAGA 724
Db 980 CCCACGAGTACCAACGCTGGACCAAGCAGAGTGAATGCCAGTTCGGGATCAGGGAGA 1039
QY 725 CCATGCCGGAACCTGGCAGGCTCT 747
Db 1040 GCCTGCCCACTCAGCAGGCTT 1062

RESULT 5
AY227205
LOCUS
DEFINITION
Rattus norvegicus CASK-interacting protein CIP98 (Cip98) mRNA, linear ROD 04-APR-2003
complete cds.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (sites)
Yap, C.C., Liang, F., Yamazaki, Y., Muto, Y., Kishida, H., Hayashida, T.,
Hashikawa, T. and Yano, R.
CIP98, a novel PDZ domain protein, is expressed in the central
nervous system and interacts with calmodulin-dependent serine
kinase
J. Neurochem. 85 (1), 123-134 (2003)
MEDLINE
22529415
PUBMED
12641734
REFERENCE
2 (bases 1 to 2763)
Yap, C.C., Liang, F., Yamazaki, Y., Muto, Y., Kishida, H., Hayashida, T.,
Hashikawa, T. and Yano, R.
Direct Submission
Submitted (29-JAN-2003) Brain Science Institute (BSI), RIKEN, 2-1
Hirose, Wako, Saitama 351-0198, Japan
FEATURES
source
1. .2763
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="WST"
/db_xref="taxon:10116"
1. .2763
/gene="Cip98"
1. .2763
/gene="Cip98"
/note="98 kDa"
/codon_start=1
/product="CASK-interacting protein CIP98"

/protein_id="AB072534.1"
/db_xref="GI:29373053"
/translation="MNAQDDGLSVSSSTGSLGAAAAAGGGGAGLRLLSANVRQLH
QALLTSPEREQYTHCLNAYHARNVFDLVRILDSVPSVKRLLPMLRLVPRRS
DQLLPDYTAEGLYLPATPYPAWAAAPDAGPGEVRLSRRARAKHEGIFGRSG
SHVGVIYVSLVEPNSLAEKGLRVGDDQILRVNDKSLARVTHAEAKLKGSKLVL
VTSAGRIPGYVTNHIVTWDPPQRSTPPSPHSGTSLRCHEDRRSAHLHLQSGDE
KKNVLVNGRSLGILIRGAEYGLGIVITVDVPGSEASGLKVDQILEVNGRSLF
SLHDAVKLLKSSHLILTVKDVGLPHARTTVDQTKWIASRIGESTINAGRPD
LIEBENKLPFGTAPAGSQVTLSSIGNQTRALLDQASHLLTEQERATMNYLDOYRG
GTISVALVMALELLNTHAKFSLSSEVRIISPDQDLRFDHLVLRRELESMAKQPP
GPGVDYTNWYSYDSTGSGTSHGTSVSSARELLMLDLMENLIDLEGTCETQGG
HSNALPDVSDVDRSPEDLPKIPPPPLAQQHDLRLGTRKPRVDRSDAPLSAA
HSGIVFSPARNRSPPPPGIAPTPTPGSPSARSPIYASISHANPSRKLPTDHL
ALVNPQIPGPFPRVOSPPGAPAEAGACACLPSPSPSEHADAMANGHFLVVEVH
RDPSEPVDNEVREALPOTRTASTLSQSDSGQTLSESDVDAGETASTSGRGTANT
KKNXKELPQTERTEGANKPGLLEPTSLIRVKSAAATLGIAIEGGANTQPLPRI
VITQRGSAHNCQQLKGVHILEVNGQTLRGKHEARAVIAEAFKTKERDIYDFLVT
EFNVML"

ORIGIN
Query Match 34.9%; Score 585; DB 10; Length 2763;
Best Local Similarity 87.8%; Pred. No. 3.4e-125;
Matches 652; Conservative 0; Mismatches 85; Indels 6; Gaps 1;
QY 5 TGGCGGCTGCCAAGGCCACGAGGGCTTGGGCTTACGATCCGTGGGGCTCGGAGCAG 64
Db 431 TGGACGCGCCAAAGCCACGAGGGCTTGGGCTTACGATCCCGGGGGCTCGGAGCAG 490
QY 65 GCGTGGGCATCTACGTGCTCTGTGGAACACAGGCTCTTAGCTGAGAGAAAGGACTGC 124
Db 491 GCGTGGGCATCTACGTGCTCTTAGTGGAGCCAGGCTCCCTGGCAGAGAGAGGTTAC 550
QY 125 GGGTGGGGACCAAGATTCCTGGCTCAACACAAATCCCTTGGCCCGGGTGACCAACGG 184
Db 551 GGGTGGGGACCAAGATTCCTGGCTCAACACAAATCCCTTGGCCCGGGTGACCAACGG 610
QY 185 AGGCGCTCAAGGCTCTGAAGGGTCCAAAGAGCTGGTGTCTGTGTACTCAGCAGGGC 244
Db 611 AGGCTGTCAGAGGCTCTCAAGGGTCCAAAGAGTGGTGTCTGTATATCTCAGCTGGC 670
QY 245 GCATCCCTGGGGCTACGTCAACCAACATCTACACCTGGTGGACCCCGAGGCGCGCA 304
Db 671 GTATACCGGGGGCTATGTCAACCAACATCTACACCTGGTGGACCCCGAGGCTGAA 730
QY 305 GCATCTCCACCTCGGGCTGCCCGAGCCCGAGCGTGGTGGCTGAGCAGCAGGAGG 364
Db 731 GCATCTCCCTCCCTCCAGGCT-----TCCCATGSCAGCACCTGAGACAGCATGAG 784
QY 365 GTGACCGGAGGAGCACCTCTGACCTCTGCAAGGAGGGATGAGAAAAGGTGAACCTGG 424
Db 785 ATGATCGAAGGAGTCCCTCACCTCTCGCAGAGTGGAGATGAGAAAAGGTGAACCTGG 844
QY 425 TGCTGGGGACGGCGGTCCTCGGCGCTCAGCATCCGTGGGGAGCTGAGTACGGCTTG 484
Db 845 TGCTGGGGACGGCGGTCCTTGGGCTCAATCCGAGGTGGAGCGAGTACGGCTTG 904
QY 485 GCATTTACATCACTGGGCTGGACCCAGGCTCTGAAGCAGAGGAGCGGGCTCAAGTTG 544
Db 905 GCATTTACATCACTGGTGTGGACCCGGGCTCTGAAGCCGAAAGCAGCGGCTCAAGTTG 964
QY 545 GGGACCAAGTCTAGAGTGAATGGGGGAGCTTCTCAACATCTCTACACGACGAGCTG 604
Db 965 GGGACCAAGTCTAGAGTGAATGGGGGAGCTTCTCAACATCTCTGATCAGCAGGAG 1024
QY 605 TCAGGCTGCTTAAGTCACTCGGCACCTCATCTGACAGTGAAGAGAGCTCGGGAGGCTGC 664
Db 1025 TGAAGTCTCTAAGTCACTCCCGGACCTCATCTGACCGTGAAGAGAGCTCGGGAGGCTGC 1084
QY 665 CCCATGCCCGACCACTGTGGACGAGACCAAGTGAATGCCAGTTCGGGATCAGGGAGA 724
Db 1085 CCCACGAGTACCAACCGTGGACCAAGTGAATGCCAGTTCGGGATCAGGGAGA 1144

QY 725 CCATGGCGAATCTGGCAGGGTCT 747
Db 1145 GCATCACCACCTCAGCAGGGTTT 1167

RESULT 6

HSMB00885

LOCUS

DEFINITION

partial cds.

ACCESSION

AL110228

VERSION

AL110228.1

KEYWORDS

Source

ORGANISM

Homo sapiens

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 2902)

Wambutt,R., Haubner,D., Meves,H.W., Gassenhuber,J. and Wiemann,S.

Direct Submission

Submitted (15-AUG-1999) MIPS, Am Klopferspitz 18a, D-82152

Martinsried, GERMANY

Cloned from S. Wiemann, Molecular Genome Analysis, German Cancer

Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;

sequenced by AGOWA (Berlin/Germany) within the CDNA sequencing

consortium of the German Genome Project.

This clone (DKFZp434N014) is available at the RZPD in Berlin.

Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059

Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further

information about the clone and the sequencing project is available

at <http://www.mips.biochem.mpg.de/proj/cdna/>.

Location/Qualifiers

1..2902

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="DKFZp434N014"

/tissue type="testis"

/clone_lib="434 (synonym: htes3). Vector pSport1; host

DH10B; sites NotI + SalI

/dev_stage="adult"

1..2902

/gene="DKFZp434N014"

<1..2173

/gene="DKFZp434N014"

/note="similarity to NV-CO-38 protein"

/codon_start=2

/product="hypothetical protein"

/protein_id="CA553685.1"

/db_xref="GI:5817167"

/db_xref="GOA:Q9UFZ3"

/db_xref="SPTREMBL:Q9UFZ3"

/translation="KLIGSRGPRLLPPECKRSVACVQALKGSKLVSVSAGRIPGGY

VNHYTWDPQCRSTSPSGLPQPHGALROEGDRSTLHLOGDEKKNVLGID

GRSLGITRGAEYGLGIYITVDGSEAEGLKVGQDILEVRSFLNLHDAVR

LUKSHRLITVDGRPHARTVDETQWASSRIRETMANAGFLGLDTTEGINKP

GFYKPGAGSQVTLSSIGNQTRVLLEQARHLNQEHEATMAYLDEYRGGSVSEALV

MAFLKLNTHAKFSLSEVRGTISPDLEFHLVLRREIESMKARQPPFGAGDTYS

MVSYSTGTSSTGSHGTSITVSARNTLDLEETGEAVQGNALPDVSDVDRSTQGL

SFKPLPRPPLAQGNLDPLGQPKLGRDLQPPSPSCSGTFVSAPQNRSPAGTA

PPGTSSAODLSSPIYASVSPANSSKEPLDAHLALVNOHPIGPBPVQSPPHLKP

SBAATVAGCLLPPSPSGPDGTGNQHFVMEVRPSEPVEVRALPQTRASTL

SQSDSGQTLSDSGVDAGEAAGPGRQSRKSELPRNETPDGANKPPG

LLPFTSLVRVKSGAATLGIALEGGAATQPLPRIVTIQRGGSAAHNCQOLKGVHVL

VNGLTLRGKEHREARIIAEAFKTDYIDFLVTFEYVNL

2824..2829

/polya_signal

/gene="DKFZp434N014"

2846

/polya_site

/gene="DKFZp434N014"

ORIGIN

Query Match 33.1%; Score 554.8; DB 9; Length 2902;

Best Local Similarity 99.6%; Pred. No. 3.5e-118;

Matches 1145; Conservative 0; Mismatches 1167; Indels 0; Gaps 0;

Matches 555; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 190 GTCAAGGCTCTGAAGGGCTCCAAAGAGCTGGTGTCTGTGTACTCAGCAGGGCGCATC 249
Db 62 GTCCAGGCTCTGAAGGGCTCCAAAGAGCTGGTGTCTGTGTACTCAGCAGGGCGCATC 121

QY 250 CTGGGGGGTACCTCAACCAACACATCTACCTGGGTGACCCCGCAGGGCGCGCAGCATC 309
Db 122 CTTGGGGGGTACCTCAACCAACACATCTACCTGGGTGACCCCGCAGGGCGCGCAGCATC 181

QY 310 TCCCCACCCCTCGGGGCTGCCCGAGCCCGGCTGGTGGTCCCTGAGGCAGCAGGAGGGTGAC 369
Db 182 TCCCCACCCCTCGGGGCTGCCCGAGCCCGGCTGGTGGTCCCTGAGGCAGCAGGAGGGTGAC 241

QY 370 CGAGGAGCACCCTCGACCTCTCTGCAAGGAGGGATGAGAAAAAGGTGAACCTGGTGTG 429
Db 242 CGAGGAGCACCCTCGACCTCTCTGCAAGGAGGGATGAGAAAAAGGTGAACCTGGTGTG 301

QY 430 GGGAGCGCCGGTCCCTGGGCGCTCAGATCCGCTGGGGGAGCTGAGTACGCGCTTGGCAT 489
Db 302 GGGAGCGCCGGTCCCTGGGCGCTCAGATCCGCTGGGGGAGCTGAGTACGCGCTTGGCAT 361

QY 490 TACATCAGCTGGCTGGACCCAGGCTCTGAAGCAGAAAGGCGGGCTCAAGGTGGGGAC 549
Db 362 TACATCAGCTGGCTGGACCCAGGCTCTGAAGCAGAAAGGCGGGCTCAAGGTGGGGAC 421

QY 550 CAGATTTAGAGTGAATGGGGAGCTTTCTCAACATCTTACACGACGAGGCTGTGACG 609
Db 422 CAGATTTAGAGTGAATGGGGAGCTTTCTCAACATCTTACACGACGAGGCTGTGACG 481

QY 610 CTGCTTAAGTCATCTCGGCACCTCATCTGACAGTGAAGGACGTCGGGAGGCTGCCCAT 669
Db 482 CTGCTTAAGTCATCTCGGCACCTCATCTGACAGTGAAGGACGTCGGGAGGCTGCCCAT 541

QY 670 GCCCGACCACTGTGACGAGCAGCAAGTGGATCGCAGTTCGCGATCAGGAGGACCATG 729
Db 542 GCCCGACCACTGTGACGAGCAGCAAGTGGATCGCAGTTCGCGATCAGGAGGACCATG 601

QY 730 GCGAATCTCGGCAGGGTCT 747
Db 602 GCGAATCTCGGCAGGGTCT 619

RESULT 7

AX714098

LOCUS

Sequence 782 from Patent EPI293569.

AX714098

ACCESSION

AX714098.1

GI:29889026

KEYWORDS

Source

ORGANISM

Homo sapiens

Homo sapiens

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1

Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S.,

Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R.,

Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagahari,K. and

Masuko,Y.

Full-length cDNAs

Patent: EP 1293569-A 782 19-MAR-2003;

Helix Research Institute (JP); Research Association for

Biotechnology (JP)

Location/Qualifiers

1..1718

/organism="Homo sapiens"

/mol_type="unassigned DNA"

/db_xref="taxon:9606"

ORIGIN

Query Match 25.1%; Score 420.2; DB 6; Length 1718;

Best Local Similarity 95.0%; Pred. No. 5.6e-87;

Matches 434; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY	4	TTGGCGCGTGCCCAAGGCCCAACGAGGGCTTTGGGCTTCAGATCCGTGGGGGCTCGGAGCAC	63
Db	1076	TTGGCGCGTGCCCAAGGCCCAACGAGGGCTTTGGGCTTCAGATCCGTGGGGGCTCGGAGCAC	1135
QY	64	GCGCTGGGCACTACCTGTCTCTGTGGTGAACGAGGCTCTTAGCTGAGAGGAAGGACTG	123
Db	1136	GCGCTGGGCACTACCTGTCTCTGTGGTGAACGAGGCTCTTAGCTGAGAGGAAGGACTG	1195
QY	124	CGGGTCGGGGACACAGATTCTTGCCTGCTCAACGACAAATCCCTGGCCCGGGTGACCCACGCG	183
Db	1196	CGGGTCGGGGACACAGATTCTTGCCTGCTCAACGACAAATCCCTGGCCCGGGTGACCCACGCG	1255
QY	184	GAGCGCGTGAAGGCTCTGAAGGGCTCGAAGAACTGGTGTCTGTCTGTACTCAGCAGGG	243
Db	1256	GAGCGCGTGAAGGCTCTGAAGGGCTCGAAGAACTGGTGTCTGTCTGTACTCAGCAGGG	1315
QY	244	CGCATCCCTGGGGGCTACGTACCAACCAATCTACACCTGGGTGGACCCGACAGGGCCCG	303
Db	1316	CGCATCTCTGGGGGCTACGTACCAACCAATCTACACCTGGGTGGACCCGACAGGGCCCG	1375
QY	304	AGCATCTCCCAACCTCGGGGCTTCGCCAGCCAGCCACGGTGGTGCCTTGAGGAGCAGGAG	363
Db	1376	AGCATCTCCCAACCTCGGGGCTTCGCCAGCCAGCCACGGTGGTGCCTTGAGGAGCAGGAG	1435
QY	364	GGTGACGGGAGGACCCCTGCACCTCTCAAGGAGGGGATGAGAAAAGGTGAACCTG	423
Db	1436	GGTGACGGGAGGACCCCTGCACCTCTCAAGGAGGGGATGAGAAAAGGTGAGTGGG	1495
QY	424	GTGCTGGGGACGGCCGGTCCCTGGGCTCACGATCC	460
Db	1496	GTGGGAAAGGAGGCCACGCTCTCAGACACCGTATTC	1532
RESULT 8			
AK056190			
LOCUS	Homo sapiens cDNA FLJ1628 fis, clone NT2R12003344, weakly similar		
DEFINITION	to PRESYNAPTIC PROTEIN SAP97.		
ACCESSION	AK056190		
VERSION	AK056190.1 GI:16551526		
KEYWORDS	oligo capping; fis (full insert sequence).		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 Ninomiya,K., Wagatsuma,M., Kanda,K., Kondo,H., Yokoi,T., Kodaira,H., Furuya,F., Takahashi,M., Kikkawa,E., Omura,Y., Abe,K., Kamihara,K., Katsuta,N., Sato,K., Tanikawa,M., Yamazaki,M., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Murakawa,K., Kaneshori,K., Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahara,K., Masuho,Y., Negai,K. and Isogai,T.		
TITLE	NEDO human cDNA sequencing project		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 1718)		
AUTHORS	Isogai,T., Otsuki,T. and Sugiyama,T.		
TITLE	Direct Submission		
JOURNAL	Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomic@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)		
COMMENT	NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.		
FEATURES	Location/Qualifiers		

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus
1 (bases 1 to 237260)
Muzry, D. Marie, Metzker, M. Lee, Abranzone, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Ayoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escoto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Greggeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C. I., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshewa, L., Louisedge, H., Lozado, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E.,
Mawhiney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K.,
Nwaokeleneh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, F., Polindexter, A., Popovic, D., Primus, E., Pu, L. L.,
Pruza, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
Rilly, B., Rilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rivers, C., Rodkey, I., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajls, D.,
Sneed, A., Sodergren, E., Song, X. Z., Sorelle, R., Sosa, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weisers, R., Smith, D. R., Holt, R. A., Smith, H. O.,
Weinstock, G. and Gibbs, R. A.
Direct Submission
Unpublished
2 (bases 1 to 237260)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (12-AUG-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 237260)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 15, 2002 this sequence version replaced gi:23664478.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated

by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GMYW
Center clone name: CH230-133E19
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 227218 bases at least Q40
Consensus quality: 229137 bases at least Q30
Consensus quality: 230338 bases at least Q20
Estimated insert size: 233753; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 235931: contig of 235931 bp in length
* 235932 236031: gap of unknown length
* 236032 237260: contig of 1229 bp in length.
----- Location/Qualifiers
1. 237260
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-133E19"
1..1635
misc_feature
/note="wgs contig"
230853..231688
misc_feature
/note="clone boundary
clone_end:T7
site:
end sequence: BH324527"
234235..235931
misc_feature
/note="wgs_end_extension
clone_end:T7"

ORIGIN
Query Match 10.7%; Score 180; DB 2; Length 237260;
Best Local Similarity 70.0%; Pred. No. 6.5e-31;
Matches 291; Conservative 0; Mismatches 110; Indels 15; Gaps 3;
QY 539 AGTTGGGACAGATTCTAGAGTGAATGGCGGAGTTTCTCAACATCTACACGACG 598
DB 220212 AGTTGGGACAGATTCTAGAGTGAATGGCGGAGTTTCTCAGCATCTCGATGACG 220153
QY 599 AGGCTGTGAGGCTGCTTAAGTCATCTCGGACCTCATCTCGACAGTGAAGGACGTCGGGA 658
DB 220152 AGGCACTGAGTGTCTCAATCTCCGCGACCTCATCTCGCGTGAAGGACGTCGGGA 220093
QY 659 GGCTCCCCCATGCCCGACCACTGTGGACGAGACCAAGTGGATGCCAGTCTCCCGATCA 718
DB 220092 GGCTGCCCGACGACGATACCACTGAGGAGTGAAGTGGATGCCAGTCTCCCGATCG 220033
QY 719 GGGAGACCATGGCGACCTCGGCGAGGTTCTGGCCATCTGTCTCGTCCAATCTCCAGACCC 778
DB 220032 GGGAAAGCATCAACCACTCAGCAGGTTCTGGCTCAGGAATGG-----GGTTCGGGATCC 219977

Qy	779	CRAGGGCCATTCTTGAAGACGAGTGNACTGC--CTCCATCCCTCCACCGCCCTGGCTC	836
Db	219976	TGGGGGGGCTCTCTGAGAGTCATTGCTCTCTCATCTGCTCCATGGGACATCCAGTCTGG	219917
Qy	837	TGCTCTCAGCCTGCAGTCCCAACACACGAGGCCCTCCATTGCGACGAGCATGACCTGGGCAC	896
Db	219916	TGTTCTCAGCCTGAGATCCAC-----TCCACTCGTGTGGTGGCACTGGTGGGC	219866
Qy	897	ATCCCTCTCTCTCTCTGGCCTCAGTTTCCCATGGAAGCTGAAATACACATCCA	952
Db	219865	CTTCCCTCTCTCTCTAGCCCAAGTGCTCTCACGAAGCTGAAGTCAGCACACACA	219810
RESULT 10			
AC127863			
LOCUS	AC127863	333531 bp	DNA linear HTG 20-NOV-2002
DEFINITION	Rattus norvegicus clone CH230-164x17, WORKING DRAFT SEQUENCE.		
ACCESSION	AC127863		
VERSION	GI:25139795		
KEYWORDS	HTG; HTGS PHASE2; HTGS DRAFT; HTGS_FULLTOP.		
SOURCE	Rattus norvegicus (Norway rat)		
ORGANISM	Rattus norvegicus		
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;		
	Rattus		
REFERENCE	1 (bases 1 to 333531)		
AUTHORS	Muzny, D., Marie, Metzker, M., Lee, A., Branzon, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlik, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howell, S., Huylk, S., Hume, J., Idlebird, P., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovac, J., Kowis, C., Kraft, C.I., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenuhewa, L., Loulseghe, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindratne, N., Mamoud, M., Mallory, K., Mangum, A., Mangun, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mathewey, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokeleneh, O., Okwuonu, G., Olarnpuaagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankoch, C., Popper, P., Poindexter, A., Popovic, D., Primus, E., Pu, L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Speed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabot, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseca, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, E., Wleczky, R., Woodson, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von		

Matches	291;	Conservative	0;	Mismatches	110;	Indels	15;	Gaps	3;
QY	539	AGTTGGGACAGATTCTAGAGTGAATGGCGGAGCTTCTCAACATCTACAGCAGC	598						
Db	178021	AGTTGGGACAGATTCTGGAGGTGAACGGCGGAGCTTCTAGCATCTGCTATGACG	178080						
QY	599	AGGCTGTGAGGCTGTTAAGTCAATCGCGACCTCATCTGACAGTGAAGACGTCGGGA	658						
Db	178081	AGGCAAGTGAAGTCTCAAGTCAATCGCGACCTCATCTGACGTAAGACGTCGGGA	178140						
QY	659	GGCTGCCCATCCCGCACCACCTGAGGAGAGACCAAGTGGATCCGATCCCGGATCA	718						
Db	178141	GGCTGCCCATCCCGCACCACCTGAGGAGAGACCAAGTGGATCCGATCCCGGATCG	178200						
QY	719	GGGAGACCATGGCAACTCGGAGGCTTGGCCACTCTGCTCGCTCCCAATCTCCAGACCC	778						
Db	178201	GGGAAAGCATCAACCACTCAGCAGGCTTGGCTCAGGAATGG----GGTTCGGGATCC	178256						
QY	779	CAGGGCCATTCTGAAGCCAGTATAGTGC--CTCCCATCTCCACCGCCCTGGGCTC	836						
Db	178257	TGGGGGGCTTCTGAGAGTCATTGCTCCTCATCTGCTCCTCATGGGACTCCCACTG	178316						
QY	837	TCCTCTCAGCTGCGAGTCCCAACACAGGGCCCTCATTTGGCAGCATGACCTGGGCAC	896						
Db	178317	TGTTCTCAGCTGAGATCCAC-----TCCACTCGTGTGGTGGCACTGGTGGGCC	178367						
QY	897	ATCCCTCTCTCTTGGGCTCAGTTTCCCATGGAAGCTGAATACACCATCCA	952						
Db	178368	CITCCCTCTCTCTCTAGCCAGTGTCTCAGAGAAGCTGAGTGCAGCACACA	178423						

RESULT 11
AL683828/c
LOCUS
DEFINITION
Mouse DNA sequence from clone RP23-340H1 on chromosome 4. Contains the 3' end of the gene for a novel collagen triple helix repeat and fibrillar collagen C-terminal domain containing protein, the Orml, Orm2 and Orm3 genes for orosomucoid 1, 2 and 3, an orosomucoid pseudogene, the gene for a novel protein similar to human AT-hook protein AKA, Gene 433437N03Rik, Gene 1110035G07Rik (C430046P22Rik) and two CpG islands, complete sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AL683828.8 GI:21212361
HTG; 1110035G07Rik; 4933437N03Rik; C430046P22Rik; collagen; CpG island; Orml; Orm2; Orm3; orosomucoid.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 214370)
Matthews, L.
Direct Submission
Submitted (17-OCT-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On May 25, 2002 this sequence version replaced gi:20218659.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep/ RP23-340H1 is from the RPI-23 Mouse PAC Library constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm VECTOR: pBACe3.6
----- Genome Center
Center: UK Medical Research Council
Center code: UK-MRC
Web site: http://mrcseq.har.mrc.ac.uk
Contact: mouseseq@har.mrc.ac.uk

This sequence is the entire insert of clone RP23-340H1. The true right end of clone RP23-6H1 is at 68592 in this sequence.

FEATURES

source
1. 214370
Location/Qualifiers
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="4"
/clone="RP23-340H1"
/clone_lib="RPCI-23"
33. 456
/notes="match: GSS: Em: AQ970708"
224. 48609
/genes="bm340H1.1"
join(-224..268,5141..5194,6819..6872,7088..7141,11429..11482,14299..14352,14778..14822,14911..14955,15176..15229,15986..16039,16652..16705,18803..18856,20304..20357,23140..23247,27044..27097,27252..27359,28291..28344,29180..29233,30946..31002,31105..31158,31602..31655,31880..31933,32539..32592,32668..32775,33068..33175,35048..35155,37066..37119,37537..37599,38049..38102,38249..38284,38390..38407,38794..38829,39742..39768,41448..41513,42153..42321,43416..43525,44900..45118,46780..46926)
/product="bm340H1.1"
/gene="bm340H1.1" (novel collagen triple helix repeat and fibrillar collagen C-terminal domain containing protein)"
/notes="match: CDNAS: Em: AB058773 Em: AK003879
match: ESTs: Em: BB205277 Em: BB526911 Em: AV231828 Em: AI449286 Em: AU120087 Em: AV595500 Em: AW652868 Em: BI557682 Em: BE683275 Em: AA296378 Em: AW436183 Em: AW658049 Em: BE756878 Em: BB621398 Em: AV600232 Em: BM545183 Em: AW446832 Em: BF773650"
/evidence="not experimental"
join(-224..268,5141..5194,6819..6872,7088..7141,11429..11482,14299..14352,14778..14822,14911..14955,15176..15229,15986..16039,16652..16705,18803..18856,20304..20357,23140..23247,27044..27097,27252..27359,28291..28344,29180..29233,30946..31002,31105..31158,31602..31655,31880..31933,32539..32592,32668..32775,33068..33175,35048..35155,37066..37119,37537..37599,38049..38102,38249..38284,38390..38407,38794..38829,39742..39768,41448..41513,42153..42321,43416..43525,44900..45118,46780..46926)
/gene="bm340H1.1"
/note="match: proteins: Tr: Q96JF7 Tr: Q9JLI2 Tr: Q96HC0 Tr: Q8UJ3 Sw: Q93484 Sw: Q9HAA3"
/codon_start=1
/evidence="not experimental"
/product="bm340H1.1 (novel collagen triple helix repeat and fibrillar collagen C-terminal domain containing protein)"
/protein_id="CAD62259.1"
/db_xref="gi:28172191"
/translation="GEPDGRPGFVGQGLMGFGLVGEPIVGEKDRGMGPPGA
PGKSGMHPGTGGIGNPGEPPGWPGSKGLPMRGAKGHRGPGDPGAGQSGK
GLKRGVGRPGPGQGAAGRGSKAGKGLGIPGSPGPAKGLPGEFGSQPGQ
VPGPPGPKPGFAGVEPOLGDSGMKGLDLPLGPPGEGLIGRCEGLGDHGP
VGPDLKGRDGPDPDGEHKGQEGKLGEGSGSPGPGITGVPGEKPGKQSGKOR
GAKGAGQGVLMGIPGEPFGTGPFGSRGLGFTGAPGFMGAGCEGFLAGYNG
HKGITGLPGPPGPKXGKQDGEDGKTGPGPDGPDGDRGDRGDPGCPGCGG

CDS

repeat_region	/note="3.0 copies 6 mer ACAAGG 27% conserved"	
repeat_region	7797..7807	
repeat_region	/note="11.0 copies 1 mer G 22% conserved"	
repeat_region	complement(8155..8227)	
repeat_region	/note="L2 repeat: matches 3221..3295 of consensus"	
repeat_region	8223..8394	
repeat_region	/gene="BM340H1.1"	
repeat_region	/note="match: GSS: Em:AZ718678"	
repeat_region	8389..8408	
repeat_region	/note="2.0 copies 10 mer CCACATCTGA 40% conserved"	
repeat_region	8579..8865	
repeat_region	/note="B4A repeat: matches 5..287 of consensus"	
repeat_region	9083..9093	
repeat_region	/note="2.2 copies 5 mer TCCTG 22% conserved"	
repeat_region	10.3%; Score 172.6; DB 10; Length 214370;	
repeat_region	Best Local Similarity 83.4%; Pred. No. 3.3e-29;	
repeat_region	Matches 196; Conservative 0; Mismatches 39; Indels 0; Gaps 0;	
QY	194	AGGCTCTGAAGGGCTCCAAAGACTGGTCTCTGTGTACTCAGCAGGCGCATCCCTG 253
DB	186488	AGGCTCTCAAGGCTCAAGAGCTGGTGTCTGTATCTCAGTGGCGTATCCCG 186429
QY	254	GGGGTCACTGTCACCAACCACTTACACCTGGGTGGACCCGAGGCGCAGCATCTCC 313
DB	186428	GGGGTATGTGACCAACCACTTACACCTGGGTGGACCCAGGGTCCGAAGCACATCC 186369
QY	314	CACCTCTGGGCTGCCAGCCCCACGGTGGTGCCTGAGGAGGAGGGTGACCGGA 373
DB	186368	CTCCCTCAGCTGCCCCAGCCCCATGGCAGCACCCTGAGCAGCGTGAAGATGCCGAA 186309
QY	374	GGAGCACCTCGCACCTCTCTGCAAGGAGGGGATGAGAAAAGGTGAACCTGGTCT 428
DB	186308	GGAGTACCTCTCCACCTCTCTGAGATGAGAAAAGGTGAGATTGACACT 186254
RESULT 12		
AF336379/c		
LOCUS	AF336379 219809 bp DNA linear HTG 02-APR-2001	
DEFINITION	Mus musculus chromosome 4 clone BAC279, *** SEQUENCING IN PROGRESS	
DEFINITION	*** 3 unordered pieces.	
ACCESSION	AF336379.1 GI:13507296	
VERSION	HTG; HTGS PHASE1.	
KEYWORDS	Mus musculus (house mouse)	
SOURCE	Mus musculus	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
REFERENCE	1 (bases 1 to 219809)	
AUTHORS	Rump,A., Varela,A., Mburu,P., Brown,S.D.M. and Rosenthal,A.	
TITLE	Mouse chromosome 4 genomic sequence	
JOURNAL	Unpublished	
REFERENCE	2 (bases 1 to 219809)	
AUTHORS	Rump,A.	
TITLE	Direct Submission	
JOURNAL	Submitted (17-JAN-2001) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstr. 11, Jena 07745, Germany	
COMMENT	* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.	
COMMENT	* This record will be updated with the finished sequence.	
COMMENT	* as soon as it is available and the accession number will be preserved.	
COMMENT	* 1 199354: contig of 199354 bp in length	
COMMENT	* 199355 199454: gap of unknown length	
COMMENT	* 199455 215338: contig of 16084 bp in length	
COMMENT	* 215339 215639: gap of unknown length	
COMMENT	* 215639 219809: contig of 4171 bp in length.	
FEATURES	Location/Qualifiers	
source	1..219809	
source	/organism="Mus musculus"	

```
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="4"
/clone="BAC279"

ORIGIN
Query Match      10.2%; Score 171.8; DB 2; Length 219809;
Best Local Similarity 87.4%; Pred. No. 5.1e-29;
Matches 188; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 539 AGGTTGGGACACAGATTCTAGAAGTGAATGGCGGAGCTTTCTCAACATCTTACACGAG 598
Db 191195 AGGTTGGAGACAGATTCTGGAGGTGAATGGCGGAGCTTTCTCAACATCTTGCATGATG 191136

QY 599 AGGCTGTGAGGCTGTTAAGTCAATCTCGGACCTCATCTCAGAGTGAAGACGTCGGGA 658
Db 191135 AGGAGTGAAGCTGTCAAGTCAATCCCGACCTCATCTCAGCGTGAAGACGTCGGAA 191076

QY 659 GCTCCCGCATGCCCGACCACTGTGGACGAGACCAAGTGAATGCGCAGTTCCCGGATCA 718
Db 191075 GCTCCCGCATGCCCGACCACTGTGGACGAGACCAAGTGAATGCGCAGTTCCCGGATCG 191016

QY 719 GGGAGACCATGCGGAATCTCGGACGGGTCTGGCCAC 753
Db 191015 GGGAAAGGTCGCCAACTCAGCAGGGTCTGGTCTC 190981

RESULT 13
AC015144/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster (fruit fly)
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 16325)
AUTHORS
Adams,M. and Venter,J.C.
DIRECT SUBMISSION
Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
This sequence was identified as CDN:10211077 by the submitter.
For further information on this sequence e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1..16325
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"

ORIGIN
Query Match      6.5%; Score 108.6; DB 2; Length 16325;
Best Local Similarity 60.8%; Pred. No. 1.6e-14;
Matches 177; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

QY 415 GTGAACCTGTGTGGGGACGCCGGTCCCTGGGCTTCACGATCCGTGGGGAGCTGAG 474
Db 14673 GTGGAGCTGTGTATCGAACTGTCTGTCTGGGCTGTATGATCGTGGCGGTGGAG 14614

QY 475 TACGGCTTGGCATTTACATCATCTGGCGTGACCCAGGCTCTGAACAGAGGACGCGG 534
Db 14613 TACGGCTGGGGATCTTTGTACCGCGGTAGATGAAGACAGCGGTGCATCGATCCGA 14554

QY 535 CTCAGGTTGGGACCAAGTTCTAGAGTGAATGGCGGAGCTTTCTCAACATCTTACAC 594
Db 14553 CTGATGTTGGGACGAGATCTTCGAGTCAATGGGCAATCTTTCTCGATGTGACGCAC 14494
```

```
QY 595 GACGAGGCTGTGAGGCTGCTTAAGTCAATCTGGGACCTCATCTCCTGACAGTGAAGGACGTC 654
Db 14493 GACGAGGCGGTGGTCAAGTGAAGTACCAACACCAAGTGGCTGCTGATACGTCGCTG 14434

QY 655 GGGAGGCTGCCCCATGCCCGACCACTGTGGACGAGACCAAGTGGATGCC 705
Db 14433 GGCAGGTCGCCCACTCTCTGCACTCCATCGAGATGGAGCCCTGGGACGCC 14383

RESULT 14
AC010699
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster (fruit fly)
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 63934)
AUTHORS
Muzny,D.M., Adams,C., Bailey,M., Barberia,J., Blankenhorn,K.,
Bodota,B., Bock,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogue,M.,
Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J.,
Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,
Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sugdang,R.,
Tabbar,P., Taylor,T., Vasquez,L., Vinson,R., Vo,O., Wabnah,M.,
Watlington,S., Weinstein,G., Weinstein,I.R., Williamson,A.,
Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D., and
Gibbs,R.
DIRECT SUBMISSION
Unpublished
2 (bases 1 to 63934)
Worley,K.C.
DIRECT SUBMISSION
Submitted (18-SEP-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Feb 18, 2000 this sequence version replaced gi:5902990.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: DRHD
Center clone name: RPC198-25L10
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; 29% of reads
Assembly program: Phrap; version 0.980611
Consensus quality: 37377 bases at least Q40
Consensus quality: 44508 bases at least Q30
Consensus quality: 49829 bases at least Q20
Estimated insert size: 47935; sum-of-contigs estimation
Quality coverage: 0.9x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 39 contigs. The true order of the pieces
* is not known and their order in this sequence record is
```

* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

1
846: contig of 846 bp in length
847: gap of unknown length
848: contig of 1607 bp in length
849: gap of unknown length
850: contig of 846 bp in length
851: gap of unknown length
852: contig of 837 bp in length
853: gap of unknown length
854: contig of 1003 bp in length
855: gap of unknown length
856: contig of 789 bp in length
857: gap of unknown length
858: contig of 839 bp in length
859: gap of unknown length
860: contig of 835 bp in length
861: gap of unknown length
862: contig of 821 bp in length
863: gap of unknown length
864: contig of 735 bp in length
865: gap of unknown length
866: contig of 838 bp in length
867: gap of unknown length
868: contig of 859 bp in length
869: gap of unknown length
870: contig of 1200 bp in length
871: gap of unknown length
872: contig of 837 bp in length
873: gap of unknown length
874: contig of 832 bp in length
875: gap of unknown length
876: contig of 833 bp in length
877: gap of unknown length
878: contig of 835 bp in length
879: gap of unknown length
880: contig of 834 bp in length
881: gap of unknown length
882: contig of 1214 bp in length
883: gap of unknown length
884: contig of 1767 bp in length
885: gap of unknown length
886: contig of 841 bp in length
887: gap of unknown length
888: contig of 825 bp in length
889: gap of unknown length
890: contig of 1295 bp in length
891: gap of unknown length
892: contig of 814 bp in length
893: gap of unknown length
894: contig of 1399 bp in length
895: gap of unknown length
896: contig of 1492 bp in length
897: gap of unknown length
898: contig of 323 bp in length
899: gap of unknown length
900: contig of 2072 bp in length
901: gap of unknown length
902: contig of 1494 bp in length
903: gap of unknown length
904: contig of 1393 bp in length
905: gap of unknown length
906: contig of 1315 bp in length
907: gap of unknown length
908: contig of 1802 bp in length
909: gap of unknown length
910: contig of 1586 bp in length
911: gap of unknown length
912: contig of 2355 bp in length
913: gap of unknown length
914: contig of 38874 bp in length

38994 42753: contig of 3860 bp in length
42754: gap of unknown length
42755: contig of 3384 bp in length
42756: gap of unknown length
42757: contig of 4262 bp in length
42758: gap of unknown length
42759: contig of 6755 bp in length
42760: gap of unknown length
42761: contig of 6700 bp in length.
42762: Location/Qualifiers
42763: 1.63934
42764: /organism="Drosophila melanogaster"
42765: /mol_type="genomic DNA"
42766: /db_xref="taxon:7227"
42767: /clone="RPC198-25L10"
42768: ORIGIN
42769: Query Match 6.5%; Score 108.6; DB 2; Length 63934;
42770: Best Local Similarity 60.8%; Pred. No. 2e-14;
42771: Matches 177; Conservative 0; Mismatches 114; Indels 0; Gaps 0;
42772: QY 415 GTGAACCTGGTCTGGGGAGCGCGGTCCTCGGCTCAGCATCCGTGGGGAGCTGAG 474
42773: DB 37329 GTGAGCTGCTGATCGAACCTGGTCACTCTCTGGGCTGATGATCCGTGGCGGTGAG 37388
42774: QY 475 TACGGCTTGGCATTTACATCACTGGCTGGAGCCAGGCTCTGAAGCAGAGGAGCGGG 534
42775: DB 37389 TACGGCTTGGCATTTGTTCACGGCTAGATAAGACAGCGTGGCAGATCGATCCGA 37448
42776: QY 535 CTCAGGTTGGGGACCAAGATTCTAGAGTGAATGGGGAGCTTCTCAACATCTACAC 594
42777: DB 37449 CTGATGATTGGCGACGAGATCCTCGAGTCAATGGGCAATCTTCTCGATGAGCGAC 37508
42778: QY 595 GACGAGCTGTCAAGCTGCTTAAGTCATCTCGGCACCTCATCTCTGACAGTGAAGAGCTC 654
42779: DB 37509 GACGAGCTGTGGTCTAGTCAAGTACCAACAGCATGTCGTGGTGATGATGAGCTG 37568
42780: QY 655 GGGAGGTCGCCATGCGCGACCACTGTGGACGAGACCAAGTGATGATCGCC 705
42781: DB 37569 GCGAAGTGGCCCACTCTGTCACCTCCATGAGATGGAGCCCTGGGACGCC 37619
42782: RESULT 15
42783: AE003536_1/c
42784: WPCOMMENT
42785: Sequence split into 4 fragments LOCUS AE003536 Accession AE003536
42786: Fragment Name Begin End
42787: AE003536_0 1 110000
42788: AE003536_1 100001 210000
42789: AE003536_2 200001 310000
42790: AE003536_3 300001 350428
42791: Continuation (2 of 4) of AE003536 from base 100001 (AE003536 Drosophila melanogaster chr
42792: Query Match 6.5%; Score 108.6; DB 3; Length 110000;
42793: Best Local Similarity 60.8%; Pred. No. 2.1e-14;
42794: Matches 177; Conservative 0; Mismatches 114; Indels 0; Gaps 0;
42795: QY 415 GTGAACCTGGTCTGGGGAGCGCGGTCCTCGGCTCAGCATCCGTGGGGAGCTGAG 474
42796: DB 47081 GTGGAGCTGCTGATCGAACCTGGTCACTCTCTGGGCTGATGATCCGTGGCGGTGAG 47022
42797: QY 475 TACGGCTTGGCATTTACATCACTGGCTGGAGCCAGGCTCTGAAGCAGAGGAGCGGG 534
42798: DB 47021 TACGGCTTGGCATTTTGTTCACGGCTAGATAAGACAGCGTGGCAGATCGATCCGA 46962
42799: QY 535 CTCAGGTTGGGGACCAAGATTCTAGAGTGAATGGGGAGCTTCTCAACATCTACAC 594
42800: DB 46961 CTGATGATTGGCGACGAGATCCTCGAGTCAATGGGCAATCTTCTCGATGAGCGAC 46902
42801: QY 595 GACGAGCTGTCAAGCTGCTTAAGTCATCTCGGCACCTCATCTGACAGTGAAGAGCTC 654
42802: DB 46901 GACGAGCTGTGGTCTAGTGAAGTACCAACAGCATGTCGTGGTGATGATGAGCTG 46842

Qy 655 GGGAGGCTGCCCCATGCCCGCACCACTGTGGACGAGACCAAGTGGATCGCC 705
Db 46841 GGCAGGTGCCCCCACTCTCTGCACTCCATCGAGATGGAGCCCTGGAGCGCC 46791

Search completed: April 22, 2004, 00:07:15
Job time : 6483.91 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 21, 2004, 21:36:07 ; Search time 3091 Seconds

(without alignments)
3645.867 Million cell updates/sec

Title: US-10-078-090-151

Perfect score: 1326

Sequence: 1 LRRAKAHEGLGFSIRGGSEH.....TWANSAGSGHSARSNLQTPG 260

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DRV=xlp
-Q=/cgn2_1/USPTO.spool.p/US10078090/runat_20042004_132810_16588/app_query.fasta_1.455
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS-bits START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORW=exc -HEAFSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10078090.cgn_1_1_5265@runat_20042004_132810_16588 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6
-FGAEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	1261.5	95.1	3566	9	AB040959 Homo sapi
2	1196.5	90.2	3178	10	AK122523 Mus muscu
3	1137.5	87.3	2763	10	AY227205 Rattus no
4	951.5	71.8	2902	9	AL110228 Homo sapi
5	706	53.2	1718	6	AX714098 Sequence
6	706	53.2	1718	9	AK056190 Homo sapi
7	438	33.0	1805	9	BC029054 Homo sapi
8	438	33.0	2037	9	AK026862 Homo sapi
9	408.5	30.8	132292	9	AL138895 Human DNA
10	408.5	30.8	171627	2	AF336382 Homo sapi
11	385.5	27.6	214370	10	AL683828 Mouse DNA
12	332	25.0	33531	2	AC127863 Rattus no
13	330.5	24.9	219809	2	AF336379 Mus muscu
14	325.5	24.5	237260	2	AC131430 Rattus no
15	321.5	24.2	2213	9	AB018687 Homo sapi
16	318.5	24.0	2162	6	AR279633 Sequence
17	318.5	24.0	2162	6	BD079792 Cancer-as
18	318.5	24.0	2162	6	AF039699 Homo sapi
19	318.5	24.0	2236	6	AR279632 Sequence
20	318.5	24.0	2236	6	BD079378 Cancer-as
21	318.5	24.0	2236	6	BD079791 Cancer-as
22	318.5	24.0	2236	9	AF039700 Homo sapi
23	318.5	24.0	2289	6	AR279636 Sequence
24	318.5	24.0	2289	6	BD079795 Cancer-as
25	318.5	24.0	2409	6	BD079637 Sequence
26	318.5	24.0	2409	6	BD079796 Cancer-as
27	315.5	23.8	2283	9	AB006955 Homo sapi
28	312.5	23.6	2063	10	AF228924 Mus muscu
29	312.5	23.6	2084	10	BC010819 Mus muscu
30	312.5	23.6	3061	10	AY103465 Mus muscu
31	312.5	23.6	3084	10	AF228925 Mus muscu
32	312	23.5	837	6	AX658639 Sequence
33	291	21.9	2184	9	BC016057 Homo sapi
34	290	21.9	16325	2	AC015144 Drosophil
35	290	21.9	63934	2	AC010699 Drosophil
36	290	21.9	110000	3	AE003536_1 Continuation (2 of
37	290	21.9	173509	3	AC010031 Drosophil
38	290	21.9	216041	2	AC112524 Drosophil
39	287	21.6	4306	9	AK024422 Homo sapi
40	265	20.0	178662	2	EX470066 Dario rer
41	265	20.0	191463	2	EX510367 Dario rer
42	265	20.0	219657	2	EX470198 Dario rer
43	261	19.7	191463	2	EX510367 Dario rer
44	252	19.0	3027	9	AK126776 Homo sapi
45	242	18.3	7234	5	BC063721 Xenopus 1

ALIGNMENTS

RESULT 1

AB040959 3566 bp mRNA linear PRI 22-FEB-2001
LOCUS Homo sapiens mRNA for KIAA1526 protein, partial cds.
DEFINITION AB040959
ACCESSION AB040959
VERSION AB040959.1 GI:7959318
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE Nagase, T., Kikuno, R., Ishikawa, K., Hirose, M. and Ohara, O.
PREDICTION OF THE CODING SEQUENCES OF UNIDENTIFIED HUMAN GENES. XVII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro
DNA Res. 7 (2), 143-150 (2000)
JOURNAL 2027482
MEDLINE 10819331
PUBMED
REFERENCE 2 (bases 1 to 3566)
AUTHORS Chara, O., Nagase, T. and Kikuno, R.
TITLE Direct Submission
JOURNAL Submitted (04-APR-2000) Osamu Ohara, Kazusa DNA Research Institute, Department of Human Gene Research; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:cdna@kazusa.or.jp, URL: http://www.kazusa.or.jp/huge, Tel: 81-438-52-3913, Fax: 81-438-52-3914)
FEATURES
source
1. .3566
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="fj04743"
/clone_lib="pBluescriptII SK plus"
1. .3566
/genes="KIAA1526"
<1. .2893
/genes="KIAA1526"
/notes="Start codon is not identified. fj04743 cDNA clone for KIAA1526 has a 1196-bp insertion after the position 1335 of the sequence of KIAA1526."
/codon_start=2
/product="KIAA1526 protein"
/protein_id="BAA96050.1"
/db_xref="GI:7959319"
/translation="SRPRPIOLYPCSPNGSSRRRLAPGPVAPVAPARTAAPPTP
RALRARPGLGVENNALDGLSVSSSTSLGSAAGAGGGGAGLLLSANVRLQAL
TALLSAEREOFTCLNAYHARNVFDLVTLEVLDSVPKRLPLMLRVLPRSDQL
LFCQYAEGLYLPATTIPKQPAWGPDSAGPEVRLVLRKAKAHEGLGFSIRGSEH
GVGIYVSLPEGLAEKGLRGVQDILRVNDLSLAKVTHAEAKVAKSKLLVLSVIS
AGRIPOGVYTHIYTVDPQGRSISPPGLPQPHGALRQEGDRRLTLHLQGDDEK
KVNLLDGRSLGTLTGAEGYGLIYITGVPDPSAEGLKVGQDQILEVNGRFLN
ILHDEAVLLKSRHLITVQVRLPHARTTVDKTIASSRIETWANSAGFLGLD
TTGINKPGPKYKAGQVTLSSLNQTRVLEQARHLLNEQEHATWAYVLDEVRG
SVSEYALVAKFLNTHAKFLLSYRGTISPDLERFCHLYLRREIESMKARPPG
FGADTSMVYSIDTSGTSGHGTITVSARLIDLETGEAVQENALPDVSVDD
VRSTQGLSKFKPLRPPLPLAQNDLPLGPKRLGREDLQPPSPMPCSGTVFSAQN
RSPAGTAPFTSSAQDLPPSPPIYASVSPANSPKRPDLNHLVNOHPIGFFPRVQ
SPHLKSPSAEATVAGCLLPSPSGHPDQTNQHFVMEVHRPDSEPDVNEVRLP
QTRASTLSQSDSGDSDGAEAGAPGRGROSSTKRSKSELPRNERPT
DGANKPGLLEPTSLVRVKKSATIGIAEGGANTRQLPLRTVIQSGSAHNCGL
KVGHVILEVNGLTLRGKEHREARIIAEAPKTKDRDYIDFLVTEFNVL"

ORIGIN
Alignment Scores:
Pred. No.: 1.43e-72 Length: 3566
Score: 1261.50 Matches: 250
Percent Similarity: 96.17% Conservative: 1
Best Local Similarity: 95.79% Mismatches: 9
Query Match: 95.14% Indels: 1
DB: 9 Gaps: 1
US-10-078-090-151 (1-260) x AB040959 (1-3566)

QY 1 LeuArgArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGlyGlySerGluHis 20
Db 596 TTGGGGGCTGCCAAGGCCACAGAGGGCTTGGGCTTCAGCATCGTGGGGGCTCGGAGCAC 655
QY 21 GlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGluGlyLeu 40
Db 656 GGGCTGGGCTATCTACGTGCTCTGTGGTGAACACAGGCTCTAGCTGAGAGGAGACTG 715
QY 41 ArgValGlyVAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAla 60
Db 716 CGGGTGGGACACAGATTCTCGCGCTCAACGACAAATCCCTGGCCGGGTGAGCCAGCGG 775
QY 61 GluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAlaGly 80
Db 776 GAGGCCCTCAAGGCTCTGAAGGGCTCAAGAGGCTGGTGTCTGTCTACTCAGCAGGG 835
QY 81 ArgIleProGlyTyrValThrAsnHisIleTyrThrTyrValAspProGlnGlyArg 100
Db 836 CGCATCTCTGGGGCTACGTCACCAACCATCTACCTGGGTGACCCGCGAGGCCCGC 895
QY 101 SerIleSerProSerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlnGlu 120
Db 896 AGCATCTCCCAACCTCGGGGCTGCCACCCAGGCTGGTGGTCCCTGAGCAGCAGGAG 955
QY 121 GlyAspArgArgSerThrLeuHisLeuLeuGlnGlyGlyVAspGluLysLysValAsnLeu 140
Db 956 GTGACCGGAGGAGCACCTCGACCTCTCAAGAGGGGATGAGAAAAGGTGAACCTG 1015
QY 141 ValLeuGlyVAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGluTyrGlyLeu 160
Db 1016 GTGTGGGGGACGCGCGGCTCCCTGGGCTCACATCGTGGGGAGCTGAGTACGGGCTT 1075
QY 161 GlyIleTyrIleThrGlyValAspProGlySerGluAlaGluGlySerGlyLeuLysVal 180
Db 1076 GGCATTTCATCATCTGGCGTGGACCCAGGCTCTGAAGCAGAGGCGGCGGCTCAAGGTT 1135
QY 181 GlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHisAspGluAla 200
Db 1136 GGGGACACAGATTCTAGAAGTGAATGGCGGAGCTTCTCAACATCTTACACGACGAGCT 1195
QY 201 ValArgLeuLysSerSerArgHisLeuLeuLeuThrValLysAspValGlyArgLeu 220
Db 1196 GTGAGGTGCTTAAGTCATCTCGGCACCTCATCTGACGTAGAGGACGTCGGGAGCTG 1255
QY 221 ProHisAlaArgThrValAspGluThrLysTrpIleAlaSerSerArgIleArgGlu 240
Db 1256 CCCCACCGCCACCATCTGGACGACCAAGTGGATCCGCGATCCCGATCAGGAG 1315
QY 241 ThrMetAlaAsnSerAlaGly---SerGlyHisSerAlaArgSerAsnLeuGlnThrPro 259
Db 1316 ACCATGCGCAACTCGGAGGCTTCTTTGGCGCATCTCAACAGAGGAATAAACAAGCCA 1375
QY 260 Gly 260
Db 1376 GGA 1378
RESULT 2
AK122523 3178 bp mRNA linear ROD 15-MAR-2003
LOCUS Mus musculus mRNA for mKIAA1526 protein.
DEFINITION AK122523
ACCESSION AK122523
VERSION AK122523.1 GI:28972777
KEYWORDS FLI_CDNA.
SOURCE Mus musculus (house mouse)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1
AUTHORS Okazaki, N., Kikuno, R., Ohara, R., Inamoto, S., Aizawa, H., Yuasa, S., Nakajima, D., Nagase, T., Ohara, O. and Koga, H.
TITLE Prediction of the coding sequences of mouse homologues of KIAA gene: II. The complete nucleotide sequences of 400 mouse KIAA-homologous cDNAs identified by screening of terminal sequences

of cDNA clones randomly sampled from size-fractionated libraries
DNA Res. 10, 35-48 (2003)
2 (bases 1 to 3178)
Okazaki, N., Kikuno, R., Nagase, T., Ohara, O. and Koga, H.
Direct Submission
Submitted (07-FEB-2003) Hisashi Koga, Kazusa DNA Research
Institute, Laboratory for Genome Informatics, 2-6-7
Kazusa-kamata, Kisarazu, Chiba 292-0818, Japan
(E-mail: mouse@kazusa.or.jp, Tel: 81-438-52-3919, Fax: 81-438-52-3918)
The CREATE program supported by Japan science and technology
corporation; cDNA full insert sequencing; Kazusa DNA Research
Institute; cDNA library construction, clone selection and 5'- &
3'-end one pass sequencing.
Location/Qualifiers
1. .3178
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="mbh03417"
/tissue_type="brain"
/dev_stage="adult"
/note="vector:modified pBC SK+."
1. .3178
/gene="mKIAA1526"
1. .2610
/gene="mKIAA1526"
/note="CDS is predicted by in silico analysis. Start codon
is not identified."
/codon_start=1
/evidence="not experimental"
/product="mKIAA1526 protein"
/protein_id="BAC65805.1"
/db_xref="GI:2897278"
translation="ANVRLHQAALTALSEPREFQTHCLNAYHARNVDFLVRLLV
LLDSPKRLPLMLVPIPRSDQLLFQYTAEGLYLPATTPYROPAAWPDGAGPGV
RLVSRRAKHEGLGFSIRGSEGVGYVSLVEPGSLAKEGLRVGDQILRVNDKSL
ARVTHAEAVKALKSLVLSVAGRIFGVITNHYTWDPGRSTSPSSLPQPH
GSTLQREDDRRSLHLSQDEKKNLVGDSRLGTLIRGAEYGLGFIITGVDPG
SEAESGLKVDQILEVNGRFLNLHDEAVKLLKSRHLTLTVGRUPHAKTIVD
QTKWASSRIGESVANSAGFPDDEGTGKPGYKGPAGSQVTLSSIGNQTRALLDD
QARHLLTQERATMAYLAQYRGVTISVEAMVMAFELLNTHAKFSLSEVRSISPO
BLDRDHLVREIESMKAROPGPGVDTYSVSYSDTSGTSGHGTSTVSSARNT
LDEGTGETQSTNALPDVSDVVDKVSDELPGIKPPPPPLAQGHDRLLGQPRK
GREDPAPLSAHSQIVSAPRNPFPPTAFTPGPSSAQDSPPSPYASISHANPS
RXLPLTHALVNHQIPFVRVQSPHLKSPPTAGACLPFPSPSEHDAVGAQ
HFVLVEHRPDEPNVNEVRLPQTRTSLQSDSGTSLSESDVDAGETEASTSGR
GROTAANKNGEQPRTAEAGAKPPGLLEPTSLTVRKSAATLIGIAIEGANT
ROPLPRIVTIQGGSAHNCGLKVGHVILEVNGQLTRGKEHKEAARIIEAFKTERD
YIDFLVTFENVNL"

ORIGIN

Alignment Scores:
Pred. No.: 2,016-68 Length: 3178
Score: 1196.50 Matches: 237
Percent Similarity: 92.72% Conservative: 5
Best Local Similarity: 90.80% Mismatches: 18
Query Match: 90.23% Indels: 1
DB: 10 Gaps: 1

US-10-078-090-151 (1-260) x AK122523 (1-3178)

Qy 1 LeuArgArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGlyGlySerGluHis 20
Db 319 CTGCGCGCCGACAGGCCACAGAGGGCTTGGGCTTCAGCATCCGCGGGCTCGGAACAC 378
Qy 21 GlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGluGlyLeu 40
Db 379 GCGGTGGGCATCTACGTGTCTCTAGTGGAGCGGGCTCCCTGGCAGAGAAGGAAGGGTTG 438
Qy 41 ArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAla 60
Db 439 CGGCTGGGGACACAGATTCCTGGCGCTCACGATAAATCTCTAGCCGGGTGACCCACGGC 498

Qy 61 GluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAlaGly 80
Db 499 GAGGCTGTCAAGGCTCTCAAGAGGCTCCAGAGGCTGGTGTCTGTATATCTACGCTGGG 558
Qy 81 ArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTrpValAspProGlnGlyArg 100
Db 559 CGTATCCAGGGGGCTATGTACCAACACATCTACACCTGGGTGGGACCCACAGGGTGA 618
Qy 101 SerIleSerProProSerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlnGlu 120
Db 619 AGCAGATCCCTCCCTCCAGCTGCCCCAGGCCCATGGCAGCACCCCTGAGACAGCGTGA 678
Qy 121 GlyAspArgArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysLysValAsnLeu 140
Db 679 GATGACCGAAGAGTAGTACCTCCACCTCTCGAGAGTGGAGATGAGAAAAGGTGAACCTG 738
Qy 141 ValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyValAlaGluTyrGlyLeu 160
Db 739 GTGTGGGGGACGGCGGCTCTTGGCCCTCAGATCCGAGGTGGAGCAGATACGSCCTT 798
Qy 161 GlyIleTyrIleThrGlyValAspProGlySerGluAlaGluGlySerGlyLeuLysVal 180
Db 799 GGCATTTTACATCACTCGTGTGGACCCAGGCTCTGAAGCAGAGAGAGCGGCGCTCAAGTT 858
Qy 181 GlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHisAspGluAla 200
Db 859 GGAGACCAAGATTCTGAGGTGAATGGCGGAGCTTCTTCAACATCTCTGCATGATGAGCA 918
Qy 201 ValArgLeuLeuLysSerSerArgHisLeuIleLeuThrValLysAspValGlyArgLeu 220
Db 919 GTGAGCTCTCTCAAGTCATCCCGCACCTCATCTTCACTGAGGACGCTCGGAAAGCTG 978
Qy 221 ProHisAlaArgThrThrValAspGluThrLysTyrIleAlaSerSerArgIleArgGlu 240
Db 979 CCCACGACGTCACACCGCTGACCAAGAGTGGATCCGAGTTCCTCCGGATCCGGGAA 1038
Qy 241 ThrMetAlaAsnSerAlaGly---SerGlyHisSerAlaArgSerAsnLeuGlnThrPro 259
Db 1039 AGGTCGCCCACTCAGAGGGTTTCCAGGGGACACACAGAGAGGAGCAAGCAAGCCCA 1098
Qy 260 Gly 260
Db 1099 GGA 1101

RESULT 3
AY227205 2763 bp mRNA linear ROD 04-APR-2003
LOCUS Rattus norvegicus CASK-interacting protein CIP98 (CIP98) mRNA,
complete cds.
DEFINITION
AY227205
ACCESSION
AY227205.1 GI:29373052
KEYWORDS
Rattus norvegicus (Norway rat)
SOURCE
Rattus norvegicus
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (sites)
Yap, C.C., Liang, F., Yamazaki, Y., Muto, Y., Kishida, H., Hayashida, T.,
Hashikawa, T. and Yano, R.
CIP98, a novel PDZ domain protein, is expressed in the central
nervous system and interacts with calmodulin-dependent serine
kinase
J. Neurochem. 85 (1), 123-134 (2003)
JOURNAL
MEDLINE
22529415
PUBMED
12641734
REFERENCE
2 (bases 1 to 2763)
Yap, C.C., Liang, F., Yamazaki, Y., Muto, Y., Kishida, H., Hayashida, T.,
Hashikawa, T. and Yano, R.
Direct Submission
Submitted (29-JAN-2003) Brain Science Institute (BSI), RIKEN, 2-1
Hirosawa, Wako, Saitama 351-0198, Japan
JOURNAL
FEATURES
Location/Qualifiers

844 GTGCTGGGGACGGCGGCTCTTGGCCCTCACATCCGAGGTGGAGCCGAGTACGCCCTT 903
161 GIVLLETYLLEThrGlyValAspProGlySerGluAlaGluGlySerGlyLeuIysVal 180
904 GGCATTTACATCACTGCTGTGGACCCGGGCTCTGAAGCCGAAAGCAGCGGCTCAAGTT 963
181 GlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHisAspGluAla 200
964 GGGGACACAGATTCTGGAGGTGAACGGGCGGAGCTTTTCAGCATCTCTGCATCAGGAGCA 1023
201 ValArgLeuLeuIysSerSerArgHisIleuIleLeuThrValIysAspValGlyArgLeu 220
1024 GTGAAGCTGCTCAAGTCATCCCGGACCTCACTCTGACCGTGAAGACCTCGGGAGGCTG 1083
221 ProHisAlaArgThrThrValAspGluThrLysTrpIleAlaSerSerArgIleArgGlu 240
1084 CCCACGACGACACACCGTGGACGACCAAGTGGATCGCCAGTTCGCGGATCGGGGAA 1143
241 ThrMetAlaAsnSerIleGly---SerGlyHisSerAlaArgSerAsnLeuGlnThrPro 259
1144 AGCATCACCACCTCAGCAGGTTTCCAGGGGACCTCACAGAGAAGGGGCAAAACAAGCCA 1203
260 Gly 260
1204 GGA 1206
RESULT 4
HSM800885 2902 bp mRNA linear PRI 18-FEB-2000
LOCUS Homo sapiens mRNA; cDNA DKFZp434N014 (from clone DKFZp434N014);
DEFINITION partial cds.
ACCESSION AL110228
VERSION AL110228.1 GI:5817166
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2902)
AUTHORS Wambutt R., Heubner, D., Mewes, H.W., Gassenhuber, J. and Wiemann, S.
TITLE Direct Submission
JOURNAL Submitted (15-AUG-1999) MIPS, Am Klopferspitze 18a, D-82152
Martinsried, Germany
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
consortium of the German Genome Project.
This clone (DKFZp434N014) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://www.mips.biochem.mpg.de/proj/cDNA/.
FEATURES
Location/Qualifiers
1..2902
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp434N014"
/tissue_type="testis"
/clone_lib="434 (synonym: htes3). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="adult"
1..2902
/gene="DKFZp434N014"
<1..2173
/gene="DKFZp434N014"
/notes="similarity to NY-CO-38 protein"
/codon_start=2
/product="hypothetical protein"
/protein_id="CA53585.1"
/db_xref="GI:5817167"
/db_xref="GOA:Q9UFZ3"
/db_xref="SPTREMBL:Q9UFZ3"

source 1..2763
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strains="WT"
/db_xref="taxon:10116"
gene 1..2763
/gene="Cip98"
1..2763
/gene="Cip98"
/note="98 kDa"
/codon_start=1
/product="CASK-interacting protein Cip98"
/protein_id="AAO72534.1"
/db_xref="GI:29373053"
/translation="MNAQLDGLSVSSSTGSLGSAAGGGGAGRLLSANVRQLH
QALTALLSEPEREQTHCLNAYHARRNFDVRLTLRLVLDSPVKRLLPMLRLVIPS
DOLLPDQYTAGLXLPAATPYROPAAAPADGAGCEVRLVSLREAKAHEGLGFSIRGG
SEHGVIYVSLVEPGLAEKGLAVGDOILRVNKSRLARVTHAEAVKALGSKVLVS
VYSKIPGVITNHIYTWDFPQKSTSPSSLPFGSTLRQEDDRKLSALHQSGBE
KKVNLVLDGSLGTLTIGGAEYGLGIYITGVDPGSEAESSGLKVGQDILLEVNGRSL
SILHDEAVKLLKSSRLTLTKVDRGLPHARTVTDQTKWIASRIGSIYNSAGFPD
LITECTNPKPGVAGQVTLSSIGNOTRALDDQARHLTEQERATMMYLYDQYRG
GTISVEALVMALFELLNTHAKFSLSEVRGIISPDLDLRFHLVLRREISMKARQP
PGVGDYTSYVSPTSGTSGHGTSTTVSSARERLLWLDLMENTLIDLEGTCTTGG
STNALPDVSDVDRSPSDLPQIKPDPPLAOGHDLGLQTKPKVREDSAPLSSAA
HSGIVTSAPRNRSPPPGCIAPTPTFPGSSARDSPSPFIYASISHAMPSSKPLDTHL
ALVNQHPFGPFRVOPSPHLPAPAGACACLPFPSPSEHADAMGANOHVFLVEVH
RPSDPDVNEVRALPQTASTLSQLSDSGQTLSEDSGDAGETASTSGRGVLTAT
KNKNGKLPQERTTEGAKPPGLLEPTSLIRVKSAAATLGAIEGANTROPPLPRI
VTIOGGGAHNCQGLKVGVILEVNGQTLRGKHEARVIAEAFKTKERYIDPLVT
EFNVML"
ORIGIN
Alignment Scores:
Pred. No.: 5,74e-66 Length: 2763
Score: 1157.50 Matches: 232
Percent Similarity: 90.80% Conservative: 5
Best Local Similarity: 88.89% Mismatches: 21
Query Match: 87.29% Indels: 3
DB: 10 Gaps: 2
US-10-078-090-151 (1-260) x AY227205 (1-2763)
QY 1 LeuArgAlaAlaHisGluGlyLeuGlyPheSerIleArgGlyGlySerGluHis 20
DB 430 CTGCGACGCGCCCAAGGCCACGAGGCTTGCGCTTCAGCATCCGCGGGGCTCGGAGCAC 489
QY 21 GlyValGlyIleTyValSerLeuValGluProGlySerLeuAlaGluGlyGlyLeu 40
DB 490 GCGCTGGGCGATCTACGTGCTCTAGTGGAGCGAGGCTCCCTGGCAGAGAAGGAGGTTA 549
QY 41 ArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAla 60
DB 550 CGGGTCGGGGACGAGATTCTGCGGCTCAACGTAATCTCTAGCCCGGGTGACCCACGCG 609
QY 61 GluAlaValIleAlaLeuIysGlySerIysLysLeuValLeuSerValTySerAlaGly 80
DB 610 GAGGCTGTCAAGGCTCTCAAGGGCTCCCAAGAGTGTGGTCTGCTGTATCTACGCTGGG 669
QY 81 ArgIleProGlyGlyTyValThrAsnHisIleTyThrTrpValAspProGlnGlyArg 100
DB 670 CGTATACCGGGGGCTATGTCTACCAACCATCTACACCTGGGTGGACCCCGAGGCTCGA 729
QY 101 SerIleSerProProSerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlu 120
DB 730 AGCACATCCCTCCCTCCAGCCTT-----CCCCATGGCAGCACCTCGAGACAGCATGAA 783
QY 121 GlyAspArgArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluIysIysValAsnLeu 140
DB 784 GATGATCGAAGGAGTCCCTTACACCTCTCTGCGAGGTGGAGATGAGAAAAGGTGAACCTG 843
QY 141 ValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyValadluTyGlyLeu 160

/translation="KLGSRGPRLLPPECRSVACVQALKSGKKLVLSVYAGRIPIGY
VTNHYTWDPQGRSISPPSLQPHGALRQEGDRSTHLLQGGDEKKVNLVLDG
RSLGLATRGAEYGLGIYITGVDPGEAGSLKVGQDILEVNGRSLFLNLHDEAVR
LLKSSRHLLTVKDVGRUPHARTTVDRTKMTASSRIETWANSAGFLDULTTEINKP
GFYKPGASOVLTLASLQNOTVLLAEQARHLNEQEHATWANYLDEYRGGSVSEALV
MALFKLNTAKFLLSEVRGTISPODLREDFHLVLEELSMKAROPPGPGAGDTYS
NYSYDTSSTGSGTSTISSAKNTLDLEETGAQGNINALEPVSVDVDRSTSQL
SSFKLPRLPPPLAGNDLPLOCPKRLGREDLQPPSPSCSGTVPSPQRNSPPAGTA
PTPCTSSAQDLSPSPITYASVPANPSSKRLDHALALVNHQPIGFPFVQSPPHLXSP
SAEATVAGCLLPSPSPQGTGTNGHFVMEVHRPDSPPDVNEVREALPQTRTASTL
SOLDSTGTLSEDSGDAGEAAGAPGRQSVSTKRSKSLPNERPDTGANKPPG
LLEPTSTAVRVKKSAAATGIALEGSANTROPLRVTIQRGSSAHNCQLKVGHVILE
VNGLTLRKEHREARIIIAEAFKTKDRDYIDFLVTEFVNL"
2824...2829
/gene="DKFZp434N014"
2846
/gene="DKFZp434N014"
polyA_signal
polyA_site
ORIGIN
Alignment Scores:
Pred. No.: 1,278-52 Length: 2902
Score: 951.50 Matches: 191
Percent Similarity: 89.95% Conservative: 6
Best Local Similarity: 87.21% Mismatches: 15
Query Match: 71.76% Indels: 7
DB: 9 Gaps: 2
US-10-078-090-151 (1-260) x HSM800895 (1-2902)
Qy 43 GlyAspGlnLeuArgValAsnAspLysSerLeuAlaArgValThrHisAlaGluAla 62
Db 20 GCGCCACGACTACTCTCCTCAGTGCGCTTCAGTGCGC-----TGT 61
Qy 63 ValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAlaGlyArgIle 82
Db 62 GTCCAGGCTCTGAAGGGCTCCAGAGCTGGTGCTGTCTGTACTCAGCAGGGCGCATC 121
Qy 83 ProGlyGlyTyrValThrAsnHisIleTyrThrTrpValAspProGlnGlyArgSerIle 102
Db 122 CCTGGGGGTACTGTCACCAACACATCTACACTCGGTGGACCCGCGAGCGCGCATC 181
Qy 103 SerProSerGlyLeuProGlnProHisGlyAlaLeuArgGlnGlnGlyAsp 122
Db 182 TCCCAACCTCGGGGCTGCCCCAGGCCACCGTGTGCCCTGAGCAGCAGAGGGTGAC 241
Qy 123 ArgArgSerThrLeuHisLeuLeuGlnGlyAspGluLysLysValAsnLeuValLeu 142
Db 242 CGGAGGAGCACCTGCACCTCTCGACGAGGGGATGAGAAAAGGTGAACCTGGTGTG 301
Qy 143 GlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGlyTyrGlyLeuGlyIle 162
Db 302 GGGGACGGCGGTCTCCCTGGGCTCAGATCCGTGGGGAGCTGAGTACGGCTTGGCATT 361
Qy 163 TyrIleThrGlyValAspProGlySerGluAlaGluGlySerGlyLeuLysValGlyAsp 182
Db 362 TACATCACTGGCGTGGACCCAGGCTCTGAAGCAGAGGAGCGGGCTCAAGGTTGGGAC 421
Qy 183 GlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHisAspGluAlaValArg 202
Db 422 CAGATTCTAGAGTGAATGGCGGAGCTTCTCAACATCTCTACAGCAGAGGCTGTCCAGG 481
Qy 203 LeuLeuLysSerSerArgHisIleLeuIleLeuThrValLysAspValGlyArgLeuProHis 222
Db 482 CTGCTTAAGTCATCTCGGCACCTCTACCTCTGACAGTGAAGAGAGCTGGGAGGTGCCCAT 541
Qy 223 AlaArgThrThrValAspGluThrLysTrpIleAlaSerSerArgIleArgGluThrMet 242
Db 542 GCCCGCACCACTGTGACGAGACCAAGTGGATCGCCAGTTCCTCCGATCAGGAGACCATG 601
Qy 243 AlaAsnSerAlaGly---SerGlyHisSerAlaArgSerAsnLeuGlnThrProGly 260
Db 602 GCGAACTCGGACGGTTCCTTGGCGATCTCACAACAGAGGAAGTAACCAAGCCAGGA 658

RESULT 5
AX714098
LOCUS
DEFINITION
Sequence 782 from Patent EPI293569.
AX714098
ACCESSION
Version
AX714098.1 GI:29889026
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
AUTHORS
Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S.,
Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R.,
Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagahari,K. and
Masuho,Y.
TITLE
Full-length cDNAs
JOURNAL
Patent: EP 1293569-A 782 19-MAR-2003;
Helix Research Institute (JP); Research Association for
Biotechnology (JP)
FEATURES
source
1..1718
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 5,378-37 Length: 1718
Score: 706.00 Matches: 140
Percent Similarity: 95.95% Conservative: 2
Best Local Similarity: 94.59% Mismatches: 6
Query Match: 53.24% Indels: 0
DB: 6 Gaps: 0
US-10-078-090-151 (1-260) x AX714098 (1-1718)
Qy 1 LeuArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGlyGlySerGluHis 20
Db 1076 TTGGCGGTGCCAAGGCCACCGAGGGCTTTGGGCTTCAGCATCCGTGGGGCTCGGAGCAC 1135
Qy 21 GlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGluGlyLeu 40
Db 1136 GGGGTGGGCATCTACGTGCTCTGGTGGAAACAGGCTCTCTAGCTGAGAAGGAGACTG 1195
Qy 41 ArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAla 60
Db 1196 CGGTCGGGACACAGATTCTGCGCGTCAACGCAAAATCCCTGGCCCGGGTGACCCACGG 1255
Qy 61 GluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAlaGly 80
Db 1256 GAGCCCGTCAAGGCTCGAAGGGCTCCAGAACTGTGTCTGTCTGTACTCAGCAGGG 1315
Qy 81 ArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTrpValAspProGlnGlyArg 100
Db 1316 CGCATCTCTGGGGCTACGTCAACCAACCATCTACCTGGGTGGACCCGCGAGGCCGC 1375
Qy 101 SerIleSerProProSerGlyLeuProGlnProHisGlyAlaLeuArgGlnGlnGlu 120
Db 1376 AGCATCTCCCACTCGGGCTCGGCCCTCCAGCCACCGCTGGTGGTCCCTGAGCAGCAGG 1435
Qy 121 GlyAspArgArgSerThrLeuHisLeuLeuGlnGlyGlyAspGlyLysLysValAsnLeu 140
Db 1436 GGTGACCGGAGGACACCTCTGCACCTCTCGAAGAGGGATGAGAAAAAGGTGAGTGGG 1495
Qy 141 ValLeuGlyAspGlyArgSerLeu 148
Db 1496 GTGGGAAAGGAGGCCAGCTCTC 1519
RESULT 6
AK056190
LOCUS
DEFINITION
Homo sapiens cDNA FLJ31628 fis, clone NT2RI2003344, weakly similar

ACCESSION	to PRESYNAPTIC PROTEIN SAP97.
VERSION	AK056190
KEYWORDS	AK056190.1 GI:16551526
SOURCE	oligo capping; file (full insert sequence).
ORGANISM	Homo sapiens (human)
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 Ninomiya, K., Wagatsuma, M., Kanda, K., Kondo, H., Yokoi, T., Kodaira, H., Furuya, T., Takahashi, M., Kikkawa, E., Omura, Y., Abe, K., Kamiyama, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yanashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuho, Y., Nagai, K. and Isogai, T.
TITLE	NEDO human cDNA sequencing project
JOURNAL	2 (bases 1 to 1718)
REFERENCE	Isogai, T., Otsuki, T. and Sugiyama, T.
AUTHORS	Direct Submission
JOURNAL	Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
COMMENT	NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.
FEATURES	Location/Qualifiers
source	1..1718
	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
	/clone="NT2R12003344"
	/cell_line="NT2"
	/cell_type="textocarcinoma"
	/clone_lib="NT2R12"
	/notes="cloning vector: pHE185FL3-mRNA from NT2 neuronal precursor cells treated 2-weeks mitotic inhibitor after 5-weeks retinoic acid (RA) induction.-majority NT2 neuron"
CDS	650..1687
	/notes="unnamed protein product"
	/codon_start=1
	/protein_id="BAB71117.1"
	/db_xref="GI:16551527"
	/translation="MNAADGLSVSSSTGSLGSAAGAGGGGAGLRLLSANVROLHQ ALTALSEAREQFTCLNAYHARNVFDLVRLLDPSVKKRLLEMLRLVIRPSD QLLFDQYTAEGLYLPATTPYQPAGWGDPSAGFVRLVSLRAKAHEGLGSLIRGS EHGVIYSLVPEGSIAEGLRVLGQILRVNDKSLARVTHAEVAKALGSKXIVLS YSAGRISGGVTHVITWVDPQGRISPSFGLPQPHGALRQOEGDRRSTLHLLQGD EKVSGVGKGGQPLRLHLLPPNEPQSCLEAARGWFCPSVFPQVCTEGWCFEAFLL FDLCSVCYNTG"
ORIGIN	
Alignment Scores:	
Pred. No.:	5.37e-37
Score:	706.00
Matches:	1718
Percent Similarity:	95.95%
Best Local Similarity:	94.59%
Query Match:	53.24%
DB:	9
US-10-078-090-151 (1-260) x AK056190 (1-1718)	
Qy	1 LeuArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGlyGlySerGluHis 20
Db	1076 TTGCGGCTGCACAGGCCAGAGGGCTTGGCTTCAGCATCCGTGGGGCTCGGAGCAC 1135
Qy	21 GlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGluGlyLeu 40
Db	1136 GGCGTGGGCATCTAGTGTCTCTGGTGAACACGAGCTCTAGCTGAGAGGAGACTG 1195
Qy	41 ArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAla 60
Db	1196 CGGGTCGGGGACACAGATTCTGGCGCTCAACGACAAATCCCTGGCCCGGTGACCCACGG 1255
Qy	61 GluAlaValIleAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAlaGly 80
Db	1256 GAGGCGCTCAAGGCTCTGAAGGGCTCCAAAGAGCTGGTGTCTGTGTACTCAGCAGGG 1315
Qy	81 ArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTrpValAspProGlnGlyArg 100
Db	1316 CGCATCTCTGGGGCTTACGTCCACCAACCATCTACCTGGGTGACCCCGGAGGCCGC 1375
Qy	101 SerIleSerProSerGlyLeuProGlnProHisGlyGlyValAlaLeuArgGlnGlu 120
Db	1376 AGCATCTCCCACTCCGGGCTTGGCCCGCCAGCCCGGTGGTGGCTGAGGACGACGAG 1435
Qy	121 GlyAspArgArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysLysValAsnLeu 140
Db	1436 GGTGACCGGAGGAGCACCTCGACCTCTGCAAGGAGGGGATGAGAAAAAGGTGAGTGG 1495
Qy	141 ValLeuGlyAspGlyArgSerLeu 148
Db	1496 GTGGGAAAGAGGAGCCAGCCTCTC 1519
RESULT 7	
BC029054	1805 bp mRNA linear PRI 07-OCT-2003
LOCUS	Homo sapiens hypothetical protein FLJ23209, mRNA (cDNA clone
DEFINITION	MGC:34150 IMAGE:5179973), complete cds.
ACCESSION	BC029054
VERSION	BC029054.1 GI:20810506
KEYWORDS	MGC.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 1805)
	Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raja, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, J., Smalish, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
TITLE	human and mouse cDNA sequences
JOURNAL	Generation and initial analysis of more than 15,000 full-length
MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
PUBMED	22388257
REFERENCE	2 (bases 1 to 1805)
AUTHORS	Strausberg, R.
TITLE	Direct Submission
JOURNAL	Submitted (01-MAY-2002) National Institutes of Health, Mammalian
	Gene Collection (MGC), Cancer Genomics Office, National Cancer
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
	USA
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov

COMMENT

Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Gunnaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louisseged, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 50 Row: a Column: 11
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 21914924.

FEATURES

Location/Qualifiers
1..1805
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:34150 IMAGE:5179973"
/tissue type="Brain, Lung, Testis, adult, pooled whole"
/clone lib="NIH MGC_115"
/lab host="DH10B"
/note="Vector: pCMV-SPORT6"
1..1805
/gene="PLJ23209"
/db_xref="LocusID:79955"
176..1711
/product="PLJ23209 protein"
/protein_id="AAH29054.1"
/db_xref="GI:20810507"
/db_xref="LocusID:79955"
/translation="MAQFVAFDFPLGLDLSLSSRLSRHLGSDSGSTATRYLL
RQQLNGPFGIRASPMGRVILINSPIEANDSEDIHSVRVPEAGRLGFSVR
GSGEHLGTFVSGKESAEAGLCVGDKITEVNGLSLETTMGSAVKVLTSSRLH
MVRVRGVPVGIKPSKEKTVVNVNRLVVEKCGSTPSDTSSEDDGVRIVHLVTTSD
DCLGNFRGGEKFGGLYVSKVDHGLAEENGKIVGQVLAANGVRPDDI:SHSQAVE
VLKQTHMLTITKETGRVPAKVEVSEYCMLDRLNSGLVQLSPASESSSVSVCSALL
AYSGSPSRDMDICLQGEPCGPGWGRADYAMQEPBAGGEVETWCSVRPTVIL
RTAIRSGPFPGRDLDSALSPKLLALLSRFPFTRISQSYLILWEKQQRKEK
SGSPGEKALQSRKTLMLNLFKGGQGRGLARDREAWLDSGSLAKTYPRLDIEKEM
GAT"
422..676
/note="PDZ; Region: Domain present in PSD-95, Dlg, and
ZO-1/2. Also called DHR (Dlg homologous region) or GLGF
(relatively well conserved tetrapeptide in these domains).
Some PDZs have been shown to bind C-terminal polypeptides"
/db_xref="CDD:smart00228"
839..1042
/note="PDZ; Region: Domain present in PSD-95, Dlg, and
ZO-1/2. Also called DHR (Dlg homologous region) or GLGF
(relatively well conserved tetrapeptide in these domains).
Some PDZs have been shown to bind C-terminal polypeptides"
/db_xref="CDD:smart00228"

ORIGIN

Alignment Scores:
Pred. No.: 1-22e-19 Length: 1805
Score: 438.00 Matches: 107
Percent Similarity: 55.13% Conservative: 38
Best Local Similarity: 40.68% Mismatches: 90
Query Match: 33.03% Indels: 28
DB: 9 Gaps: 6

US-10-078-090-151 (1-260) x BC029054 (1-1805)

Qy 1 LeuArgAlaLysAlaHisGluGly--LeuGlyPheSerIleArgGlyGlySerGlu 19

Db	434	GTCCGGGTGGAGAGAGTCCAGCAGGAGGCTTACGCTGCGGGGGTCTCAGAG	493
Qy	20	HisGlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGluGly	39
Db	494	CATGCCCTGGGCACTTCTGTCAGCAAGTGGAGAGCAGCAGTGCAGAGCGGGCTGGC	553
Qy	40	LeuArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHis	59
Db	554	CTGTGCGTGGGGACACAGATCAGGAGGTGAATGGCTGAGCTGGAGAGCACCACCATG	613
Qy	60	AlaGluAlaValLysAlaLeuLysGlySerLysLysValLeuValLeuSerValTyrSerAla	79
Db	614	GGTAGCGCGGTAAAGGTGTCACAGCAGCAGCCCTGCACATGATGGTTCGGCGCATG	673
Qy	80	GlyArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTyrValAspProGlnGly	99
Db	674	GGCGGTGTCCGGCATCAAGTCTCCAGGAGAGACACGCTGGTGGATGTGTGAAT	733
Qy	100	ArgSerIle-----SerProSerGlyLeuProGlnProHisGly	113
Db	734	CGGCGCTGTGTAGTGAGAGAGTGCCTTCAACACCTCCGAC-----	775
Qy	114	GlyAlaLeuArgGlnGlnGluGlyAspArgArgSerThrLeuHisLeuLeuGlnGly	133
Db	776	-----ACCAGCTCAGAGATGTTGTCGGCGC---ATCGTCCACCTATACACACCTCC	826
Qy	134	AspGluLysLysValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArg	153
Db	827	GACGAC-----TTCTGCTGGGCTTCAACATCCGT	856
Qy	154	GlyGlyAlaGluTyrGlyLeuGlyIleTyrIleThrGlyValAspProGlySerGluAla	173
Db	857	GGGGCAAGAGTTGGCTGGCATCTATGTCTCCAAAGTGGACCATGTTGGCTGGCC	916
Qy	174	GluGlySerGlyLeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeu	193
Db	917	GAGGAGAATGGCATCAAGTGGGGACCAAGTCTCTGGCAGCCACGCTGCAGGTTGAC	976
Qy	194	AsnIleLeuHisAspGluAlaValArgLeuLeuLysSerSerArgHisLeuIleLeuThr	213
Db	977	GACATCAGCCACAGCAGCGCGTGGAGTGTGAAGGGCCAAACGACATCATGCTGACC	1036
Qy	214	ValIysAspValGlyArgLeuProHisAlaArgThrValAspGluThrIleTyrIle	233
Db	1037	ATCAAGGAGACCGCGCGTATCTCTGCTACAGAGAGATGTTTCTGAGTACTGCTGGCTG	1096
Qy	234	-----AlaSerSerArgIleArgGluThrMetAlaAsnSerAlaGlySerGlyHisSer	251
Db	1097	GACGACTGAGCAACGGGGTGTGCAGACGCTGTCCCGGCCCTCTGAGAGCAGCTCCAGC	1156
Qy	252	AlaArgSer 254	
Db	1157	GTCTCTTCG 1165	
RESULT 8			
AK026862			
LOCUS	Homo sapiens cDNA: FLJ23209 fis, clone ADH00512.	2037 bp	linear mRNA
DEFINITION			
ACCESSION	AK026862		
VERSION	AK026862.1	GI:10439820	
KEYWORDS	oligo capping; fis (full insert sequence).		
SOURCE	Homo sapiens (human)		
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			

AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.
 TITLE Direct Submission
 JOURNAL Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan [E-mail:flcdna@iims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416]
 COMMENT NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- and 3'-end one pass sequencing; Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).
 FEATURES
 source Location/Qualifiers
 1..2037 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="ADSH00512"
 /tissue_type="adipose tissue"
 /clone_lib="ad"
 /note="cloning vector pME183FL3"
 216..1789 /note="unnamed protein product"
 /codon_start=1
 /protein_id="BAB15577.1"
 /db_xref="GI:10439821"
 /translation="NAQGFVGFDPGLGLDSSGLSSLSRHHGSDSGSTATRYLL RQQLLLNGPPRGIPASPMGVILLINSPIANSDESDIHSRVKESFAGLGFVSVR GSEHGLGFVSKVEGSSAERAGLCVDKITEVNGLSLETMTGSAVKVLTSSRLH MVRMRGVRFGIKFKEKTKTVDVNVNRLIVKESGTPSDTSSEGGVRRIVHLTYTSD DFLGNFRIGKFEGLGIVKVDHGLAEANGIKVGDQVLAANGVRDDIHSQAVE VIKGQTHIMLTIKETGRYPAYKEMVSEYCWLDRLNSGLQSPASSESSVSCAS APVSGSLPSDRMDICLGOEPEGRPGMRADTMQTEPDAGGRVETWCSVRPTVIL RTATRSQDPHPRRLDSALSPTKALLALSRPRPTIRTSQSYLTWEKQQRKKE KSGSPKQKALQSKTLNLFYKGGKQGLADGRREATLDSGLAKTYPRLDIEKE MGVSPCCPG"
 CDS
 Alignment Scores:
 Pred. No.: 1,39e-19 Length: 2037
 Score: 438.00 Matches: 107
 Percent Similarity: 55.13% Conservative: 38
 Best Local Similarity: 40.68% Mismatches: 90
 Query Match: 33.03% Indels: 28
 DB: 9 Gaps: 6
 US-10-078-090-151 (1-260) x AK026862 (1-2037)
 QY 1 LeuArgAlaLysAlaHisGluGly---LeuGlyPheSerIleArgGlyGlySerGlu 19
 Db 474 GTCCGGTGGAGAAGAGTCCAGCAGGAGCTGGCTTTAGGCTGCGCGGGGCTCAGAG 533
 QY 20 HisGlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGlyGly 39
 Db 534 CATGGCTGGGCATCTTCGTGACAAAGTGGAGAGGAGCAGTGCAGAGCGGGTGGC 593
 QY 40 LeuArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHis 59
 Db 594 CTGTGGTGGGGACAAAGTACAGGAGGTGAATGGGCTGAGCTGAGAGCACCACCATG 653
 QY 60 AlaGluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAla 79
 Db 654 GGTAGGCGCGTAAAGGTGCTGACAGCAGCAGCGCGCTGCACATGATGTTGGCGCATG 713
 QY 80 GlyArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTyrValAspProGlnGly 99
 Db 714 GGCCGTCGTCGGGCATCAAGTCTCTCCAGGAGAACACACGAGGGGTGATGTGTGAAT 773
 QY 100 ArgSerIle-----SerProSerGlyLeuProGlnProHisGly 113

Db 774 CGCGCGCTGGTAGTGAGAGTGGTTCACACACCTCCGAC----- 815
 QY 114 GlyAlaLeuArgGlnGlnGluGlyAspArgArgSerThrLeuHisLeuLeuGlnGly 133
 Db 816 -----ACCAGCTCAGAGATGGTGTCCGGC-----ATCGTCCACCTATACACACCTCC 866
 QY 134 AspGluLysLysValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArg 153
 Db 867 GACGAC-----TTCTGCTGGGCTTCAACATCCGT 896
 QY 154 GlyGlyAlaGluTyrGlyLeuGlyIleTyrIleThrGlyValAspProGlySerGluAla 173
 Db 897 GGGGGCAAGAGTTCCTGGCATGTATGTCTCAAAAGTGGACCATGGTGGGTGGCC 956
 QY 174 GluGlySerGlyLeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeu 193
 Db 957 GAGGAGATGGCATCAGGTAGGGACGAGTCTGGCAGCCACGCTGCAGGTTCAGTTGAC 1016
 QY 194 AsnIleLeuHisAspGluAlaValArgLeuLeuLysSerSerArgHisLeuLeuLeuThr 213
 Db 1017 GACATCAGCCACAGCCAGCGCTGGAGTGTCTGAAGGGCCAAACGACCATCATGCTGACC 1076
 QY 214 ValLysAspValGlyArgLeuProHisAlaArgThrValAspGluThrIleTyrPile 233
 Db 1077 ATCAGGAGACCGCGCGTATCTGCTACAGGAGATGTTTCTGAGTACTGCTGCTG 1136
 QY 234 -----AlaSerSerArgIleArgGluThrMetAlaAsnSerAlaGlySerGlyHisSer 251
 Db 1137 GACCGACTGACAAACGGGGTGTGTGCAGCAGCTGTCCCGGCTCTGAGAGCAGCTCCAGC 1196
 QY 252 AlaArgSer 254
 Db 1197 GTCTCTTCG 1205
 RESULT 9
 AL138895/c Human DNA sequence from clone Rp11-9M16 on chromosome 9, complete
 LOCUS AL138895
 DEFINITION Human DNA sequence from clone Rp11-9M16 on chromosome 9, complete sequence.
 ACCESSION AL138895
 VERSION AL138895.16 GI:14575083
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 132292)
 AUTHORS Skuce,C.
 TITLE Direct Submission
 JOURNAL Submitted (27-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk requests: clonerequest@sanger.ac.uk
 COMMENT On Jun 28, 2001 this sequence version replaced gi:14456172. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em; EMBL; Sw; SWISSPROT; Tr; TREMBL; Wp; WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping

Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr9>
 RP11-9M16 is from the library RPCI-11.1 constructed by the group of
 Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
 VECTOR: pBACe3.6

IMPORTANT: This sequence is not the entire insert of clone
 RP11-9M16 it may be shorter because we sequence overlapping
 sections only once, except for a 100 base overlap.
 The true right end of clone RP11-9M16 is at 132292 in this
 sequence. The true left end of clone RP11-402G3 is at 124695 in
 this sequence. The true right end of clone RP11-8211 is at 2000 in
 this sequence.

FEATURES

source

Location/Qualifiers

1..132292
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="9"
 /clone="RP11-9M16"
 /clone_lib="RPCI-11.1"

repeat_region
 1..32
 /note="AluY repeat: matches 1..32 of consensus"
 repeat_region
 74..301
 /note="L1MC1 repeat: matches 5718..5955 of consensus"
 repeat_region
 399..696
 /note="Alusq repeat: matches 1..298 of consensus"
 repeat_region
 1186..1273
 /note="MIR repeat: matches 49..142 of consensus"
 repeat_region
 2271..2435
 /note="MER63A repeat: matches 46..206 of consensus"
 repeat_region
 2467..2553
 /note="MIR repeat: matches 79..165 of consensus"
 repeat_region
 2587..2713
 /note="L2 repeat: matches 2622..2740 of consensus"
 repeat_region
 2911..3207
 /note="AluSc repeat: matches 5..301 of consensus"
 repeat_region
 3214..3262
 /note="MIR repeat: matches 115..166 of consensus"
 repeat_region
 4193..4262
 /note="MIR repeat: matches 60..130 of consensus"
 4368..5308
 /note="CPG island"
 misc_feature
 evidence=not_experimental
 5299..5606
 /note="AluX repeat: matches 12..312 of consensus"
 repeat_region
 5686..5787
 /note="MIR repeat: matches 141..262 of consensus"
 repeat_region
 6714..6799
 /note="MIR repeat: matches 20..114 of consensus"
 repeat_region
 6995..7437
 /note="MLT1D repeat: matches 3..503 of consensus"
 repeat_region
 7440..7502
 /note="MIR repeat: matches 82..144 of consensus"
 repeat_region
 8034..8179
 /note="LTR16A repeat: matches 200..354 of consensus"
 repeat_region
 8160..8271
 /note="LTR16C repeat: matches 255..369 of consensus"
 repeat_region
 10037..10131
 /note="MIR repeat: matches 14..123 of consensus"
 repeat_region
 10604..10711
 /note="L2 repeat: matches 2405..2521 of consensus"
 repeat_region
 11151..11327
 /note="MIR repeat: matches 18..205 of consensus"
 repeat_region
 11579..11883
 /note="AluX repeat: matches 1..304 of consensus"
 repeat_region
 11886..12024
 /note="MIR repeat: matches 97..243 of consensus"
 repeat_region
 12598..12717
 /note="MIR repeat: matches 52..181 of consensus"
 repeat_region
 13839..14002
 /note="MIR repeat: matches 80..258 of consensus"
 repeat_region
 14071..14205

/note="MIR repeat: matches 92..226 of consensus"
 15216..15653
 /note="MLT1CB repeat: matches 1..492 of consensus"
 15702..15995
 /note="L2 repeat: matches 2372..2688 of consensus"
 16113..16847
 /note="L2 repeat: matches 1940..2707 of consensus"
 16875..16922
 /note="12 copies 4 mer atga 81% conserved"
 17482..17573
 /note="MIR repeat: matches 125..230 of consensus"
 18163..18379
 /note="MIR repeat: matches 23..241 of consensus"
 18972..19080
 /note="MIR repeat: matches 34..144 of consensus"
 20542..20934
 /note="Charlie4 repeat: matches 26..506 of consensus"
 21421..21521
 /note="L2 repeat: matches 2045..2158 of consensus"
 21523..21758
 /note="AluJo repeat: matches 1..232 of consensus"
 22863..23243
 /note="MSTA repeat: matches 1..426 of consensus"
 23319..23495
 /note="MIR repeat: matches 5..192 of consensus"
 24438..24542
 /note="MIR repeat: matches 28..140 of consensus"
 24993..25526
 /note="L2 repeat: matches 2099..2709 of consensus"
 25630..25867
 /note="MIR repeat: matches 16..262 of consensus"
 25889..26233
 /note="L2 repeat: matches 2363..2674 of consensus"
 26257..26381
 /note="MIR repeat: matches 28..153 of consensus"
 26919..27000
 /note="MIR repeat: matches 64..147 of consensus"
 27292..27435
 /note="MIR repeat: matches 102..254 of consensus"
 29060..29358
 /note="AluX repeat: matches 1..298 of consensus"
 29523..29664
 /note="L2 repeat: matches 2593..2734 of consensus"
 29669..29777
 /note="MIR repeat: matches 17..140 of consensus"
 30134..30243
 /note="L2 repeat: matches 2167..2289 of consensus"
 30460..30627
 /note="MER5B repeat: matches 1..173 of consensus"
 30658..30770
 /note="L2 repeat: matches 2572..2690 of consensus"
 31123..31357
 /note="L2 repeat: matches 2514..2731 of consensus"
 32095..32170
 /note="19 copies 4 mer acac 88% conserved"
 32365..32396
 /note="16 copies 2 mer tc 84% conserved"
 32426..32485
 /note="MIR repeat: matches 96..152 of consensus"
 33041..33127
 /note="MIR repeat: matches 20..115 of consensus"
 33494..33764
 /note="AluJo repeat: matches 4..281 of consensus"
 34352..34417
 /note="L2 repeat: matches 2641..2709 of consensus"
 34554..34848
 /note="AluX repeat: matches 1..293 of consensus"
 34946..35142
 /note="MIR repeat: matches 8..207 of consensus"
 35436..35738
 /note="AluX repeat: matches 1..303 of consensus"
 35834..35965
 /note="MIR repeat: matches 2..147 of consensus"

repeat_region 36769..36954
/note="MER33 repeat: matches 1..188 of consensus"
repeat_region 37658..37805
/note="MER repeat: matches 40..190 of consensus"
repeat_region 37865..37962
/note="MER repeat: matches 49..140 of consensus"
repeat_region 39025..39247
/note="MER repeat: matches 14..252 of consensus"
repeat_region 39184..39263
/note="12 repeat: matches 2671..2750 of consensus"
repeat_region 39516..39569
/note="MER repeat: matches 98..154 of consensus"
repeat_region 40503..40600
/note="MER repeat: matches 30..126 of consensus"
repeat_region 40730..40829
/note="MER repeat: matches 46..146 of consensus"
repeat_region 40854..40893

Alignment Scores:
Pred. No.: 1,11e-15 Length: 132292
Score: 408.50 Matches: 90
Percent Similarity: 70.55% Conservatives: 13
Best Local Similarity: 61.64% Mismatches: 32
Query Match: 30.81% Indels: 11
DB: 9 Gaps: 2

US-10-078-090-151 (1-260) x AL138895 (1-132292)

QY 3 ArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGlyGlySerGluHisGlyVal 22
Db 85432 AAATCGAGGCTAGAGAGGAAAGGGA-----CTTCAGGGCTGG----- 85394

QY 23 GlyIleTyrValSerLeuValGluProGlySerLeuAlaGluGlyGluGlyVal 42
Db 85393 -----CCTCTTCTTCCATGCTGGGCTCTGAGAAAACCAAGGCTGGGCCAG 85346

QY 43 GlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAlaGluAla 62
Db 85345 GGAGGGGTCTCTCCACACCCCGCACCTCCCTTCTTAACCTTTGTATCCCTCTCT 85286

QY 63 ValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAlaGlyArgIle 82
Db 85285 CCTAGGCTCTGAAGGCTCCAGAGAGCTGGTCTCTGTGTACTCAGAGGGCGCATC 85226

QY 83 ProGlyGlyTyrValThrAsnHisIleTyrThrTyrValAspProGlnGlyArgSerIle 102
Db 85225 CCTGGGGGCTAGTCCACCAACCACTACACCTGGTGGACCCGAGGGCGGCAGCATC 85166

QY 103 SerProSerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlnGlyAsp 122
Db 85165 TCCCAACCTCGGCTCCGCCACCCAGCCCGCTGGTGGTGGTGGTGGTGGTGGTGGTGG 85106

QY 123 ArgArgSerThrLeuHisLeuGlnGlyGlyAspGlnLysValAsnLeuValLeu 142
Db 85105 CGGAGGAGCCCTCGACCTCTCTGCAAGAGGGGATGAGAAAAGGTGAGTGGGTGGGG 85046

QY 143 GlyAspGlyArgSerLeu 148
Db 85045 AAAGGAGCCAGCCTCTC 85028

RESULT 10
AF336382/c 171627 bp DNA linear HTG 02-APR-2001
LOCUS Homo sapiens chromosome 9 clone R-3183P23 map 9q33, *** SEQUENCING
DEFINITION IN PROGRESS ***, 3 unordered pieces.
ACCESSION AF336382
VERSION AF336382.1 GI:13507299
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 171627)

AUTHORS Rump,A., Varela,A., Mburu,P., Brown,S.D.M. and Rosenthal,A.
TITLE Mouse chromosome 4 genomic sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 171627)
AUTHORS Rump,A.
JOURNAL Direct Submission
COMMENT Submitted (17-JAN-2001) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
source
1. 171627
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="9"
/map="9q33"
/clone="R-3183P23"

ORIGIN
Alignment Scores:
Pred. No.: 1.48e-15 Length: 171627
Score: 408.50 Matches: 90
Percent Similarity: 70.55% Conservatives: 13
Best Local Similarity: 61.64% Mismatches: 32
Query Match: 30.81% Indels: 11
DB: 9 Gaps: 2

US-10-078-090-151 (1-260) x AF336382 (1-171627)

QY 3 ArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGlyGlySerGluHisGlyVal 22
Db 154223 AAATCGAGGCTAGAGAGGAAAGGGA-----CTTCAGGGCTGG----- 154185

QY 23 GlyIleTyrValSerLeuValGluProGlySerLeuAlaGluGlyGluGlyVal 42
Db 154184 -----CCTCTTCTTCCATGCTGGGCTCTGAGAAAACCAAGGCTCTGGGCCAG 154137

QY 43 GlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAlaGluAla 62
Db 154136 GGGAGGGGTTCCTCCACACCCCGGCACCTCCCTTCTCTAACCTTTGTATCCCTCTCT 154077

QY 63 ValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAlaGlyArgIle 82
Db 154076 CCTAGGCTCTGAAGGCTCCAGAGAGCTGGTGTGTGTGTACTCAGAGGGCGCATC 154017

QY 83 ProGlyGlyTyrValThrAsnHisIleTyrThrTyrValAspProGlnGlyArgSerIle 102
Db 154016 CCTGGGGGCTAGTCCACCAACCACTACACCTGGTGGACCCGAGGGCGGCAGCATC 153957

QY 103 SerProSerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlnGlyAsp 122
Db 153956 TCCCAACCTCGGCTCCGCCACCCAGCCCGCTGGTGGTGGTGGTGGTGGTGGTGGTGG 153897

QY 123 ArgArgSerThrLeuHisLeuGlnGlyGlyAspGlnLysValAsnLeuValLeu 142
Db 153896 CGGAGGAGCCCTCGACCTCTCTGCAAGAGGGGATGAGAAAAGGTGAGTGGGTGGGG 153837

QY 143 GlyAspGlyArgSerLeu 148
Db 153836 AAAGGAGCCAGCCTCTC 153819

RESULT 11
AL683828/c
LOCUS
DEFINITION
AL683828
Mouse DNA sequence from clone RP23-340H1 on chromosome 4 Contains
the 3' end of the gene for a novel collagen triple helix repeat and
fibrillar collagen C-terminal domain containing protein, the Ormi,
Orm2 and Orm3 genes for orosomucoid 1, 2 and 3, an orosomucoid
pseudogene, the gene for a novel protein similar to human AT-hook
protein AKNA, gene 4933437N03Rik, gene 1110035007Rik
(C430046P22Rik) and two CpG islands, complete sequence.
ACCESSION
AL683828.8 GI:21212361
VERSION
HTG; 1110035007Rik; 4933437N03Rik; C430046P22Rik; collagen; CpG
island; Ormi; Orm2; Orm3; orosomucoid.
KEYWORDS
Mus musculus (house mouse)
SOURCE
Mus musculus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 214370)
REFERENCE
Matthews, L.
Direct Submission
Submitted (17-OCT-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On May 25, 2002 this sequence version replaced gi:20218659.
COMMENT
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clones, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw.,
SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP
database can be found at:
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-340H1 is
from the RP23-340 Mouse PAC library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6
----- Genome Center
Center: UK Medical Research Council
Center code: UK-MRC
Web site: http://mrcseq.har.mrc.ac.uk
Contact: mouseq@har.mrc.ac.uk

This sequence is the entire insert of clone RP23-340H1. The true
right end of clone RP23-6H1 is at 68592 in this sequence.
FEATURES
source
Location/Qualifiers
1..214370
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="4"
/clone="RP23-340H1"
/clone_lib="RPCI-23"
33..456
/note="match: GSS; Em:AQ970708"
224..48609
/gene="bm340H1.1"
join<224..268,5141..5194,6819..6872,7088..7141,
11429..11482,14299..14352,14778..14822,14911..14955,
15176..15229,15986..16039,16652..16705,18803..18856,
20304..20357,23140..23247,27044..27097,27252..27359,
28291..28344,29180..29233,30946..31002,31105..31158,
31602..31655,31880..31933,32539..32592,32668..32775,
33068..33175,35048..35155,37066..37119,37537..37599,
38049..38102,38249..38284,38390..38407,38794..38829,
39742..39768,41448..41513,42153..42321,43416..43525,
44900..45118,45780..46926)
/gene="bm340H1.1"
/note="match: proteins: Tr:Q96JF7 Tr:Q9JLI2 Tr:Q96HCO
Tr:Q8UUJ3 Sw:Q93484 Sw:Q9HAA3"
/codon_start=1
/evidence=not_experimental
/product="bm340H1.1 (novel collagen triple helix repeat
and fibrillar collagen C-terminal domain containing
protein)"
/protein_id="CAD62259.1"
/db_xref="GI:28172131"
/translation="GEPDGRPGPVGEQGLMGFGLVGGPGVGEKDRGVMPGGA
PGRKSGMHPETPGICNGPEPWPGRGRLPGVRCAKGRGPGDPGPAQEQSG
GLKRVPRPGPGQQAAGRGHSGAGFLGIPGFPSPGPKAKGLRFGSGQPGP
PVGPGMGKPGPGVAGFELPDGMDGLGLPGFPGGLIGRGEPELEDHGP
VGDPLKGDGPDGDEHGEKQEGKGLGDSGPGPGITGVPRGEPKQGEKGR
GAKGAKHGQVLGEMGIPGPGPGTGPGRGRTGLFTGAPRMAQAGPFLAGYNG
HKGITPLGPGPKGEGKDGKTEGPPGPDGPGVGDGRGPGPGVPGQBG
VQGRGPGQGGPGHPRGRGPKSGKEGPKGPKGAGPSRGRTGQGLPGP
RGVVRGPETAGSDIIPDRGRPGVQGGQNDGDPGVPAGRRGNPAGVLPGAQ
GPPFGESGLFGQLGPPRGEGGLFGNGEPGSGKQPGDPSGEMGPGVAGLFG
PKPGPGIGFGKIQGPPGLMGKGLIIPQDDLGAAFTQMDAQAVRSEGYPPQLALDQGG
RGPPRGRPGPPGPPHPIQFQDDLGAAFTQMDAQAVRSEGYPPQLALDQGG
BIFKTLVLSNLIQSIKTLPTKENPARVCDLMDCRQADGTWYDPLNGSSDTI
EVSCLFTGGOTCLKPTASKAEFAVRVOMFLHLLSSGTHITICLMTVWQEG
PGRSRAQVRFRAWGVREAGQGRPEVSMDCCKVHDGRWHQTLFTFTQDQQLP
IVSNDLPVSSGQRYLEVGPAFL"
repeat_region
/note="CYR11 MM repeat: matches 401..438 of consensus"
repeat_region
/note="CYR11 MM repeat: matches 1..226 of consensus"
repeat_region
/note="CYR11 MM repeat: matches 148..401 of consensus"
repeat_region
/note="8.3 copies 6 mer GGGAGA 73% conserved"
repeat_region
/note="2.9 copies 8 mer GAGCTTCA 37% conserved"
repeat_region
/note="1684..1836)
/note="Charlie4 repeat: matches 21..170 of consensus"
repeat_region
/note="2.0 copies 6 mer TTCCTG 24% conserved"
repeat_region
/note="MIR repeat: matches 10..248 of consensus"
repeat_region
/note="L2 repeat: matches 3233..3310 of consensus"
repeat_region
/note="B1F repeat: matches 25..148 of consensus"
misc_feature
/gene="bm340H1.1"

YU, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G. and Gibbs, R.A.
Direct Submission
Unpublished
2 (bases 1 to 333531)
Worley, K.C.
Direct Submission
Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 333531)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 20, 2002 this sequence version replaced GI:22855918.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GZ2N
Center clone name: CH230-164M17
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 204590 bases at least Q40
Consensus quality: 205853 bases at least Q30
Consensus quality: 206819 bases at least Q20
Estimated insert size: 208567; sum-of-contigs estimation
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 333531: contig of 333531 bp in length.

FEATURES
source
1. .333531
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-164M17"
1. .1223
/note="wgs contig"
misc_feature
213898..215752
/note="wgs contig"
misc_feature
332494..333531
/note="wgs contig"
ORIGIN
Alignment Scores:

Pred. No.: 2,73e-10 Length: 333531
Score: 332.00 Matches: 81
Percent Similarity: 57.86% Conservative: 11
Best Local Similarity: 50.94% Mismatches: 39
Query Match: 25.04% Indels: 28
DB: 5
Gaps: 5
US-10-078-090-151 (1-260) x AC127863 (1-333531)
QY 62 AlaValLysAlaLeuLysGlySerLysLeuValLeuSerValTySerAlaGlyArg 81
Db 141154 TCTTTCAGGCTCTCAAGGGCTCCCAAGAGTGGTGTCTGTATACCTGCGGCT 141213
QY 82 IleProGlyGlyTyThrValThrAsnHisIleTyThrTpValAspProGlnGlyArgSer 101
Db 141214 ATACCCGGGGGTATGTCTACCAACACATCTACACCTGGTGTGACCCCGAGGTCGAAGC 141273
QY 102 IleSerProProSerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlnGly 121
Db 141274 ACATCCCTCCCTCCAGCCTT-----CCCATGGCAGCACCTCGAGACGATGAAGAT 141327
QY 122 AspArgArgSerThrLeuHisIleLeuGlnGlyGlyAspGlyLysValAsnLeuVal 141
Db 141328 GATCGAAGGAGTGCCCTACACCTCTTCAGAGTGGAGATGAGAAAGGTGAGA----- 141381
QY 142 LeuGlyAsp-----GlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGlyTyr 158
Db 141382 TCTGGCCAGTTTAAATGGTAAAGATTTGGGGCCCTGG----- 141420
QY 159 GlyLeuGlyIleTyThrIleThrGlyValAspProGlySerGluAlaGluGlySerGlyLeu 178
Db 141421 -----GGATGCTTGGTTTCAAGCTCTCTTG 141447
QY 179 LysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHisAsp 198
Db 141448 TATACCCAGGAGCTCCCTCAAGCTAGAGGAGGCG-----GAGCTCCCTGAC 141498
QY 199 GluAlaValArgLeuLeuLysSerSerArgHisLeuLeuLeuThrValLysAspVal 217
Db 141499 CAGGCTCTATTACTTCTTTTGGTCTCTCAAGTTTGGTATTCCTGCTTCTCTCTG 141555
RESULT 13
AF336379/c 219809 bp DNA linear HTG 02-APR-2001
LOCUS Mus musculus chromosome 4 clone BAC279, *** SEQUENCING IN PROGRESS
DEFINITION *** 3 unordered pieces.
ACCESSION AF336379
VERSION AF336379.1 GI:13507296
KEYWORDS HTG; HTGS PHASE1.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 219809)
Rump, A., Varela, A., Mburu, P., Brown, S.D.M. and Rosenthal, A.
Unpublished
2 (bases 1 to 219809)
Rump, A.
Direct Submission
Submitted (17-JAN-2001) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 199354: contig of 199354 bp in length
* 199355 199454: gap of unknown length
* 199455 215538: contig of 16084 bp in length

Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hognes, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Joliviet, A., Karwathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowicz, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Loresunhewa, L., Louised, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munitadasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaekaleme, O., Okwuonu, G., Olarunpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C., Plopper, F., Polindexter, A., Popovic, D., Primus, E., Fu, L., L., Puazo, M., Qulroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shivartsbeyn, A., Sison, I., Sitter, C. D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steinitz, M., Strong, R., Sutton, A., Svatek, A., Tabors, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasantana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlarczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,

Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaekelemeh, O., Okwunou, G., Olarumpesagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankoch, C., Plopper, F., Polindexter, A., Popovic, D., Primus, E., Pu, L., Puazo, M., Quarzo, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Renly, E., Reuter, M., Richards, S., Riggs, F.,

Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,

Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,

Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,

Niednerhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
 Weinstock, G. and Gibbs, R.A.
 Direct Submission

2 (bases 1 to 237260)
Rat Genome Sequencing Consortium.
Direct Submission

Submitted (22-Nov-2002); Human Genome Sequencing Center,
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 237260)

Direct Submission
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One

On Nov 15, 2002 this sequence version replaced gi:23664478. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas

in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sentence contexts are ordered and oriented and separated

by sized gaps filled with Ns to the estimated size and sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome

shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
 Center: Baylor College of Medicine

Center code: BCM
Web site: www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
Project Information

```

Center project name: GMYW
Center clone name: CH230-133E19
----- Summary Statistics -----
Assembly: program: Phrap, version: 0.990329

```

Consensus quality: 227218 bases at least Q40
Consensus quality: 229137 bases at least Q30
Consensus quality: 230338 bases at least Q20

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

```

-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 235931: contig of 235931 bp in length
* 235932 236031: gap of unknown length
* 236032 237260: contig of 1229 bp in length.
*
FEATURES             Location/Qualifiers
     source            1..237260
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-133E19"
     misc_feature      1..1635
                     /notes="wgs contig"
     misc_feature      230853..231688
                     /notes="clone boundary"
                     clone_end:T7
     misc_feature      234235..235931
                     end sequence: BH324527"
                     /note="wgs_end_extension"
                     clone_end:T7"
ORIGIN
Alignment Scores:
Pred. No.:          4,94e-10          Length:          237260
Score:              325.50            Matches:          67
Percent Similarity: 89.29%            Conservative:     8
Best Local Similarity: 79.76%          Mismatches:      8
Query Match:        24.55%            Indels:          1
DB:                  2                Gaps:             1

US-10-078-090-151 (1-260) x AC131430 (1-237260)
QY 167 ValAspProGlySerGluAlaGluGlySerGly---LeuLysValGlyAspGlnIleLeu 185
Db 220252 CTGATCAGGTACGCCATCGTACAGTCTGGTCTGGCAGGTGGGACGACGATTCG 220193

186 GluValAsnGlyArgSerPheLeuAsnIleLeuHieAspGluAlaValArgLeuLys 205
Db 220192 GAGGTGAACGGCGGAGCTTTCAGCATCTCGATGACGAGGAGTGAAGTCTCAAG 220133

206 SerSerArgHisIleuLeuThrValLysAspValGlyArgLeuProHisAlaArgThr 225
Db 220132 TCATCCCGGACCATCTCCTGACCGTGAAGAGCTCGGAGGCTGCCCGACGACGTACC 220073

226 ThrValAspGluThrLysTrpIleAlaSerSerArgIleArgGluThrMetAlaAsnSer 245
Db 220072 ACCGTGACCCAGACCAAGTGTATCCCGGATCCCGGATCCCGGAAAGCATCACCACTCA 220013

246 AlaGlySerGly 249
Db 220012 GCAGGGTCTGGC 220001

RESULT 15
LOCUS AB018687
DEFINITION Homo sapiens mRNA for autoimmune enteropathy-related antigen
AIE-75, complete cds.
ACCESSION AB018687
VERSION AB018687.1 GI:5231270
KEYWORDS autoimmune enteropathy-related antigen AIE-75; AIE-75.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Kobayashi, I., Imamura, K., Kubota, M., Ishikawa, S., Yamada, M.,
Itonoki, H., Okano, M., Storch, W.F., Moriuchi, T., Sakiyama, Y. and
Kobayashi, K.
Identification of an autoimmune enteropathy-related 75-kilodalton
antigen
Gastroenterology 117 (4), 823-830 (1999)
99431904
MEDLINE
PUBMED 10500064
2 (bases 1 to 2213)
Kobayashi, I.
Direct Submission
Submitted (14-OCT-1998) Ichihiro Kobayashi, Hokkaido University
School of Medicine, Department of Pediatrics; North-15, West-7,
Kita-ku, Sapporo, Hokkaido 060-8638, Japan
(E-mail: ichikobae@med.hokudai.ac.jp, Tel: 011-716-1161,
Fax: 011-736-9287)
Location/Qualifiers
     1..2213
     /organism="Homo sapiens"
     /mol_type="mRNA"
     /db_xref="taxon:9606"
     /tissue_type="small intestine"
     90..1748
     /codon_start=1
     /product="autoimmune enteropathy-related antigen AIE-75"
     /protein_id="BAAB1740.1"
     /translation="MDRKVAREFRHKVDFLIENDAEKDLYDLVRMYHOTMDVAVLVG
DLKLVINEPSRLPLFDALIPKLGHOVEDQLTTPRRSKLKEVRLDLRHLPEGLGLSV
RGLEFCGGLPIHLIKGGQADSLQGVQDEIVIRNGISSTCTHEVILRTKTV
SIKVHIGLIPVKSSDEPTMQYVDFVSGVSGVSGVSGVSGVSGVSGVSGVSGVSGV
SRGLGCSISGIVAKPGIFISHVKPGLSRAEVLGEGDQIVENVGDFNSLDHKEAVN
VLKSRSLTISIVAAGRELFTMDRELRLAEQRLQRLQRLQRLQRLQRLQRLQRLQRLQ
QEMERQRKEIAQAAEENRYRKEMQIIVEEKEFKQWKEQWKEQWKEQWKEQWKEQWKE
EVHPVLRPKYDQGVPELEPADLDGGTEQEQDFRYEEGDFYSPFTPEQMG
KDVELLRKKEGSLDLALGEGVDSPIGKVVVSAVYERGAARHGIGVKGDEINAIKNG
IVTDYTLAEAEAAALQKAWNGGDWIDLVAVCPPEYDELTFE"
     2213
     /note="8 a nucleotides"

polyA_site
ORIGIN
Alignment Scores:
Pred. No.:          5,21e-12          Length:          2213
Score:              321.50            Matches:          94
Percent Similarity: 51.98%            Conservative:     44
Best Local Similarity: 35.34%          Mismatches:     102
Query Match:        24.25%            Indels:          27
DB:                  9                Gaps:             8

US-10-078-090-151 (1-260) x AB018687 (1-2213)
QY 1 LeuArgArgAlaLysAlaHis---GluGlyLeuGlyPheSerIleAtgGlyGlySerGlu 19
Db 351 GTGCGTCTGGACCGTCTGCACCCGAGCGCTCGGCTCGGCTGAGTGTGGTGGCTGGAG 410

20 HisGlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGly 39
Db 411 TTTGGCTGGGCTCTTCATCTCCACCTCATCAAGGCGCTCAGGACAGACGCTGGG 470

40 LeuArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHis 59
Db 471 CTCACGAGTGGGACGAGATCGTCGAGATCAATGATATTCATCTCTCTGTACCATC 530

60 AlaGluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAla 79
Db 531 GAGGAGGTTCATCACTCATTCGAACCAAGAAACCTGCTCATCAAGTGCAGACATC 590

80 GlyA-GlleProGlyGlyTyrvValThrAsnHisIleTyThrTrp-----ValAspPro 97
Db 591 GGCTGATCCCGTGAAGAGCTCTCTGATGAGCCCTCACTTGGCAGTATGTGAT--- 647

```

```

QY 98 GlnGlyArgSerIleSerProProSerGlyLeuProGlnProHisGlyGlyAlaLeuArg 117
Db 648 -----CAGTTTGTCTCGGAATCTGGGGCGTGGAGCGAGCCCTGGGC----- 689
QY 118 GlnGlnGluGlyAspArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysLys 137
Db 690 ---TCCCTCTGGAATCGGAAAC-----AAGGAGAGAGAG 722
QY 138 ValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGlu 157
Db 723 GTCTTCATCAGCTGTGTAGGCTCCGAGGCTTGGCTGCAGCATTTCCAGCGGCCCATC 782
QY 158 TyrGlyLeuGlyIleThrGlyValAspProGlySerGluAlaGluGlySerGly 177
Db 783 CAGAACCTGGCATCTTTATCAGCCATGTAAACCTGGCTCCCTGCTGCTGAGGTGGGA 842
QY 178 LeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHis 197
Db 843 TTGGAGATAGGGACCAAGATTGTGAGTCAATGGCGTGCACCTTCTTAACCTGGATCAC 902
QY 198 AspGluAlaValArgLeuLeuLysSerSerArgHisLeuIleLeuThr---VallysAsp 216
Db 903 AAGGAGGCTGTAATGTGTGAAGAGTAGCCGAGCTGACCATCTCCATTGTAGTGCA 962
QY 217 ValGlyArgLeuProHisAlaArgThrThrValAspGluThrLysTrpIleAlaSerSer 236
Db 963 GCTGGCCGG-----GAGCTGTTTCATGACAGACCGGAGCGGCTGGCAGAGGGG 1010
QY 237 ArgIleArgGluThrMetAlaAsnSerAlaGlySerGlyHisSerAlaArgSerAsnLeu 256
Db 1011 CGGCAGCGTGAGCT-GCAGCGGCGAGGAGCTTCTCATGCAGAACGGGCTGGCGATGGAGTC 1069
QY 257 Gln-----ThrProGly 260
Db 1070 CAACAAGATCCTCCAGGA 1087

```

Search completed: April 22, 2004, 03:01:40
Job time : 3322 secs

CA321445 UI-M-FW0-
AL049078 DKF2p434B
BF969269 602269794
CB209310 AGENCOURT
BF953476 RC3-NN118
BF953475 RC3-NN118
BF953545 RC3-NN118
BF953545 RC3-NN118
BF953479 RC3-NN118
BF953480 RC3-NN118
BF953472 RC3-NN118
BF953542 RC3-NN118
BF953470 RC3-NN118
BU166723 AGENCOURT
BB872069 BB872069
CA327501 UI-M-FY0-
BF953522 RC3-NN118
CB693399 AMGNNUC:S
BF953546 RC3-NN118
CF661317 CGLM09a34
BF953532 RC3-NN118
CD750225 AGENCOURT
CD310081 StrPu691.
CA949304 It4B01.Y
BX880758 BX880758
BX885444 BX885444
BX868932 BX868932
BG62377 602796268
AW74249 UP58409.Y
BF074654 22211.NA
BM940881 UI-M-CG0P
CB692138 AMGNNUC:T
CB773455 AMGNNUC:T
AZ625547 IMC465B12
AA938876 op74b06.s
BF737050 FMI-KT004
AL287778 Tetradon
BF953481 RC3-NN118
BZ217432 CH230-250
BH288018 CH230-62F
BH290228 CH230-114
AL272891 Tetradon

CA321445 729 14
AL049078 432 9
BF969269 693 10
CB209310 904 14
BF953476 325 10
BF953475 324 10
BF953545 324 10
BF953545 342 10
BF953479 342 10
BF953480 308 10
BF953472 342 10
BF953542 342 10
BF953470 342 10
BU166723 1174 13
BB872069 383 10
CA327501 675 14
BF953522 310 10
CB693399 343 14
BF953546 362 10
CF661317 1160 14
BF953532 334 10
CD750225 870 14
CD310081 643 14
CA949304 267 13
BX880758 713 13
BX885444 736 13
BX868932 761 13
BG62377 799 12
AW74249 534 10
BF074654 531 10
BM940881 586 12
CB692138 374 14
CB773455 396 14
AZ625547 641 28
AA938876 305 9
BF737050 342 10
AL287778 1045 29
BF953481 147 10
BZ217432 485 28
BH288018 730 28
BH290228 665 28
AL272891 961 29

5 438.4 26.1 729 14
6 432 25.8 432 9
7 409.8 24.4 693 10
8 337.4 20.1 904 14
9 308.4 18.4 325 10
10 305.4 18.2 324 10
11 301.8 18.0 342 10
12 298.6 17.8 342 10
13 293 17.5 308 10
14 293 17.5 342 10
15 292 17.4 342 10
16 282.8 16.9 342 10
17 272 16.2 1174 13
18 284.8 15.8 383 10
19 255.6 15.2 675 14
20 254.8 15.2 310 10
21 247.6 14.8 343 14
22 244.6 14.6 362 10
23 220 13.1 1160 14
24 212.6 12.7 334 10
25 205.6 12.5 870 14
26 209.4 12.5 643 14
27 206 12.3 267 13
28 188.8 11.3 713 13
29 188.8 11.3 736 13
30 188.8 11.3 761 13
31 186 11.1 799 12
32 185.6 11.1 534 10
33 184.2 11.0 531 10
34 178 10.6 586 12
35 172.8 10.3 374 14
36 172.8 10.3 396 14
37 168 10.0 641 28
38 159.2 9.5 305 9
39 130.4 7.8 342 10
40 123.2 7.3 1045 29
41 117.4 7.0 147 10
42 117.2 7.0 485 28
43 117.2 7.0 730 28
44 115.6 6.9 665 28
45 112 6.7 961 29

ALIGNMENTS

RESULT 1
BC014524
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

BC014524 4569 bp mRNA linear HTC 19-NOV-2003
IMAGE:3834205, with apparent retained intron.
BC014524
BC014524.1 GI:17939549
HTC.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 4569)
Straussberg,R.L., Feingold,E.A., Grouse,L.H., Shenmen,C.M., Schuler,G.D.,
Klausner,R.D., Collins,F.S., Wagner,L., Buetow,K.H., Schaefer,C.P., Bhat,N.K.,
Alschul,S.F., Zeeberg,B., Buetow,K.H., Moore,T., Max,S.I., Wang,J., Heish,F.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Heish,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Schaefer,C.P., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
Kernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalobon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,

OM nucleic - nucleic search, using sw model
Run on: April 21, 2004, 21:05:04 ; Search time 4115.23 Seconds
(without alignments)
12169.150 Million cell updates/sec
Title: US-10-078-090-48
Perfect score: 1677
Sequence: 1 gaggtagcggtgccaagc.....aagcgagaccgggggggaga 1677
Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55026578
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database :
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pin:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mus:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rnd:*
26: em_gss_pbg:*
27: em_gss_vrt:*
28: gb_gss1:*
29: gb_gss2:*
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
SUMMARIES
Result No. Score Query Match Length DB ID Description
1 742.4 44.3 4569 11 BC014524 Homo sapi
2 602 35.9 602 10 BE990168
3 531.8 31.7 905 13 BX390092
4 471.8 28.1 651 12 B1732824
B1732824 603353931

Db 262 GGTGGGCACTACGTGTCTTAGTGGAGCCGGGCTCCCTGGCAGAGAAGAAAGGGTTC 321
QY 125 GGGTCGGGACAGATTCTGGCGCTCAACGCAAAATCCCTGGCCGGGTGACCCACGCGG 184
Db 322 GGGTCGGGACAGATTCTGGCGCTCAACGCAAAATCTTAGCCGGGTGACCCACGCGG 381
QY 185 AGCCGCTCAAGGCTCTGAAGGCTCAAGAGAGCTGGTCTGTGTGTACTCAGCAGGGC 244
Db 382 AGGCTGTCAAGGCTCTCAAGGCTCAAGAGAGCTGGTCTGTGTGTACTCAGCAGGGC 441
QY 245 GCATCCCTGGGGCTACGTCACCAACCAATCTACACCTGGGTGGACCCGCGAGGCCGCA 304
Db 442 GATATCCAGGGGCTATGTGACCAACCAATCTACACCTGGGTGGACCCGCGAGGGTGA 501
QY 305 GCATCTCCCAACCTGGGCTGGCCCGCTCCGACCTGCAAGGAGGGATGAGAAAGGTGAACCTGG 424
Db 502 GCACATCCCTCCCTCCAGCTGGCCCGCTCCGACCTGCAAGGAGGGATGAGAAAGGTGAACCTGG 621
QY 365 GTGACCGGAGGAGCACCCTGCACTCTGCAAGGAGGGATGAGAAAGGTGAACCTGG 424
Db 562 ATGACCGAAGGAGTACCTCTCACTCTGCAAGGAGGGATGAGAAAGGTGAACCTGG 621
QY 425 TGTGGGGGACCGCGGCTCCCTGGGCTCACGATCGTGGGGAGCTGAGTACGCGCTTG 484
Db 622 TGTGGGGGAGCGCGGCTCTGGGCTCACGATCGGAGTGGAGCAGAGTACGCGCTTG 681
QY 485 GCATTACATCATGCGTGGAGCCAGGCTCTGAGCAGAGGAGCG 532
Db 682 GCATTACATCATGCGTGGAGCCAGGCTCTGAGCAGAGGAGCG 729

RESULT 6
AL049078 432 bp mRNA linear EST 04-SEP-2003
LOCUS DKFP434B1519.r1.434 (synonym: htes3) Homo sapiens cDNA clone
DEFINITION DKFP434B1519, mRNA sequence.

ACCESSION AL049078
VERSION AL049078.1 GI:4728387
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE EST (Ottewaelder, et al.)
AUTHORS Ottewaelder, B., Obermaier, B., Newes, H.W., Gassenhuber, J. and Wiemann, S.
TITLE Unpublished (1999)
JOURNAL Contact: MIPS
COMMENT MIPS

Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by MediGenomix (Martinsried/Germany) within the cDNA sequencing consortium of the German Genome Project. No 51 sequence available.
This clone (DKFP434B1519) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES
source
1..432
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFP434B1519"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="434 (synonym: htes3)"
/note="Vector: pSport1; Site_1: NotI; Site_2: SalI"

ORIGIN

Query Match 25.8%; Score 432; DB 9; Length 432;
Best Local Similarity 100.0%; Pred. No. 4e-74; 0; Gaps 0;
Matches 432; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 883 CATGACCTGGGACATCCCTCTCTCTGGCTCAGTTTCCCATGGAAGCTGAAAT 942
Db 1 CATGACCTGGGACATCCCTCTCTCTGGCTCAGTTTCCCATGGAAGCTGAAAT 60
QY 943 ACACCATCCACTGCTCTCATTTTATTTGTCCTCCCAATTAATTAATCTATCTATAGAC 1002
Db 61 ACACCATCCACTGCTCTCATTTTATTTGTCCTCCCAATTAATTAATCTATCTATAGAC 120
QY 1003 CTTAGTTGCTTCAATCAAAAAGTGGGACCATAAACCTGCCCTCATCCAGATCTGTGCA 1062
Db 121 CTTAGTTGCTTCAATCAAAAAGTGGGACCATAAACCTGCCCTCATCCAGATCTGTGCA 180
QY 1063 GATGAAAGAGAGGAGGAGGAGGAGAGAGAGATGCTTTGGGGTGTATTTGGCCAGAG 1122
Db 181 GATGAAAGAGAGGAGGAGGAGGAGGAGAGAGATGCTTTGGGGTGTATTTGGCCAGAG 240
QY 1123 GCCACCAAGGCTGGATCCCATGAAAGAAATCTGGGTGAGAGGGTCTTAAAGTCATAAACTGA 1182
Db 241 GCCACCAAGGCTGGATCCCATGAAAGAAATCTGGGTGAGAGGGTCTTAAAGTCATAAACTGA 300
QY 1183 GATCCAGTGGCAGGTGGCTGCATAGTGGCAACAGTGTATGTGCACCTTTTGTCTT 1242
Db 301 GATCCAGTGGCAGGTGGCTGCATAGTGGCAACAGTGTATGTGCACCTTTTGTCTT 360
QY 1243 CATCAGAAATCTCAGGCTGGTGGCCACCTGGCCAAATACACTGCAGAGCATGTGTCTG 1302
Db 361 CATCAGAAATCTCAGGCTGGTGGCCACCTGGCCAAATACACTGCAGAGCATGTGTCTG 420
QY 1303 TCTGTCTGTCTG 1314
Db 421 TCTGTCTGTCTG 432

RESULT 7
BF969269 693 bp mRNA linear EST 22-JAN-2001
LOCUS 602269794F1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4358158 5',
DEFINITION mRNA sequence.

ACCESSION BF969269
VERSION BF969269.1 GI:12336484
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE NIH-MGC <http://mgi.nci.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Invitrogen, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at:
<http://image.llnl.gov>
Plate: LLAM995 row: k column: 23
High quality sequence stop: 643.

FEATURES
source
1..693
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4358158"
/tissue_type="adrenal cortex carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_84"

/note="Organ: adrenal gland; Vector: pCMV-SPT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; Oligo-dr primed. Average insert size 1.229 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

ORIGIN

Query Match 24.4%; Score 409.8; DB 10; Length 693;
Best Local Similarity 95.0%; Pred. No. 98-70;
Matches 434; Conservative 0; Mismatches 1; Gaps 1;
QY 4 TTGGCGGCTGCCAAGGCCACAGAGGCTTTGGGCTTCAGCATCCGTTGGGGCTTCGAGAC 63
DB 36 TTGGCGGCTGCCAAGGCCACAGAGGCTTTGGGCTTCAGCATCCGTTGGGGCTTCGAGAC 95
QY 64 GGGCTGGGATCTACGTGTCTGTGGAAACAGGCTCTCTAGCTGAGAAAGGACTG 123
DB 96 GGGCTGGGATCTACGTGTCTGTGGAAACAGGCTCTCTAGCTGAGAAAGGACTG 155
QY 124 CGGGCTGGGACACAGATTTCTGGCGCTCAACGCAAAATCCCTGGCCGGGTGACCCACGC 183
DB 156 CGGGCTGGGACACAGATTTCTGGCGCTCAACGCAAAATCCCTGGCCGGGTGACCCACGC 215
QY 184 GAGCGCTCAAGGCTCTGAAGGCTCCAAAGAGCTGTGTCTGTGTACTCAGCAGG 243
DB 216 GAGCGCTCAAGGCTCTGAAGGCTCCAAAGAGCTGTGTCTGTGTACTCAGCAGG 275
QY 244 CGCATCCCTGGGGCTACGTCACCAACCATCTACACCTGGGTGACCCGAGGCGCGC 303
DB 276 CGCATCCCTGGGGCTACGTCACCAACCATCTACACCTGGGTGACCCGAGCGCA 334
QY 304 AGCATCTCCCAACCTGGGCTGCCCCAGCCAGCCAGCGTGGTCCCTGAGGAGGAG 363
DB 335 AGCATCTCCCAACCTGGGCTGCCCCAGCCAGCCAGCGTGGTCCCTGAGGAGGAG 394
QY 364 GGTGACCGGAGGACCCCTGACCTCTCTCAAGGAGGGATGAGAAAAGGTGAACCTG 423
DB 395 GGTGACCGGAGGACCCCTGACCTCTCTCAAGGAGGGATGAGAAAAGGTGAGTGG 454
QY 424 GTGCTGGGGACGCGCGGTCCCTGGGCTCAGCATCC 460
DB 455 GTGGGAAAGAGGACCCCTCTCAGACACCGTATTC 491

RESULT 8
CB209310 904 bp mRNA linear EST 05-FEB-2003
LOCUS
DEFINITION
AGENCOURT 11348497 NIH_MGC_164 Mus musculus cDNA clone
IMAGE:30243230 5', mRNA sequence.

ACCESSION
CB209310
VERSION
CB209310.1 GI:28250873
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)

REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 904)
NTH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)

COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. David Rowe and Dr. Mina
cDNA Library Preparation: Invitrogen Corp
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDAM0314 row: c column: 15
High quality sequence start: 18
High quality sequence stop: 487.
Location/Qualifiers

FEATURES

1. 904
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30243230"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_164"
/notes="Vector: pCMV-SPT6; Site_1: EcoRV; Site_2: NotI;
Non-normalized full-length enriched library from pooled
mouse embryonic limb, maxilla and mandible, day 10.5 and
11.5 (size selected for the 0.5-1 kb fragments) Cloned
directionally, priming method: Oligo-dr. cDNA enrichment:
5'k bp. Average insert size 1.8k bp. Priming sequence:
5'GACGTAGTTAGATCGGAGCGCGCCCTTT 3'. Tissue contributed
by David Rowe. Library constructed by ResGen, Invitrogen
Corp."

ORIGIN

Query Match 20.1%; Score 337.4; DB 14; Length 904;
Best Local Similarity 87.8%; Pred. No. 13e-55;
Matches 368; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
QY 129 CGGGGACAGATTCTGGCGCTCAACGCAAAATCCCTGGCCGGGTGACCCACGCGAGGC 188
DB 21 CGGGATCAGATTCTGGCGCTCAACGCAAAATCTCTAGCCCGGTGACCCACGCGAGGC 80
QY 189 CGTCAAGGCTCTGAAGGCTCCAAAGAGCTGTGTCTGTGTACTCAGCAGGCGCAT 248
DB 81 TGTCAAGGCTCTCAAGGCTCCAAAGAGCTGTGTGTGTATCTCAGCTGGGGGTAT 140
QY 249 CCCTGGGGGCTACGTCAACCAACCATCTACCTGGGTGACCCGCGGCGCAGCAT 308
DB 141 CCCAGGGGCTATGTGACCAACCATCTACCTGGGTGACCCACGAGGTGAGACAC 200
QY 309 CTCGCCACCTCGGGCTGCCCCAGCCAGCGTGGTGGCTGAGGAGGAGGTGA 369
DB 201 ATCCCTCTCCCTCAGCGCTGCCAGCCCATGCGAGCAGCCCTGAGACAGGTGAAGATGA 260
QY 369 CGGAGGAGGACCTGACCTCTCTGAGGAGGGATGAGAAAAGGTGAACCTGGTGT 428
DB 261 CGAAGGAGTACCTCTCCCTCTGAGAGTGGAGATGAGAAAAGGTGAACCTGGTGT 320
QY 429 GGGGAGCGCGGCTCCCTGGGCTCAGCATCCGTTGGGAGCTGAGTACGGCTTTGGCAT 488
DB 321 GGGGAGCGCGGCTCCCTGGGCTCAGCATCCGAGTGGAGAGTACGGCTTTGGCAT 380
QY 489 TTATCATCTGCGCTGGAGCCAGGCTCTGAGCAGAGGAGCGGCTCAGGTTGGG 547
DB 381 TTATCATCTGCGTGGAGCCAGGCTCTGAGCAGAGAGCGGCTCAGGTTGATGG 439

RESULT 9

BF953476/c
LOCUS
DEFINITION
RC3-NN182-121100-011-g08 NN182 Homo sapiens cDNA, mRNA sequence.
ACCESSION
BF953476
VERSION
BF953476.1 GI:12370751
KEYWORDS
EST.
SOURCE
Homo sapiens (human)

ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 (bases 1 to 325)
Dias Neto, S., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V.,
O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.

TITLE
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE
20202663

10737800
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?ti=RC3&t2=RC3-NN1182-
121100-011-908&t3=2000-11-12&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 325.

FEATURES
source
1..325
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="NN1182"
/note="Organ: nervous normal; Vector: puc18; Site: 1: Smal;
Site 2: Smal; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

ORIGIN
Query Match 18.4%; Score 308.4; DB 10; Length 325;
Best Local Similarity 99.7%; Pred. No. 68-50; Indels 0; Gaps 0;
Matches 309; Conservative 0; Mismatches 1;
QY 375 GAGCACCCCTGACCTCTCTCAAGAGGGGATGAGAAAAGTGAACCTGGTCTGGGGA 434
DB 321 GAGCACCCCTGACCTCTCTCAAGAGGGGATGAGAAAAGTGAACCTGGTCTGGGGA 262
QY 435 CGGCCGCTCTGGGCTACGATCCGTGGGGAGCTGAGTACGGCTTGGCATTTACAT 494
DB 261 CGGCCGCTCTGGGCTACGATCCGTGGGGAGCTGAGTACGGCTTGGCATTTACAT 202
QY 495 CACTGGGTGGACCCAGGCTCTGAAGCAGAAGGACGCGGCTCAAGGTTGGGGACAGAT 554
DB 201 CACTGGGTGGACCCAGGCTCTGAAGCAGAAGGACGCGGCTCAAGGTTGGGGACAGAT 142
QY 555 TCTAGAAGTGAATGGGGAGGCTTTCTCAACATCTACAGCAGGCTGTCAAGGCTGCT 614
DB 141 TCTAGAAGTGAATGGGGAGGCTTTCTCAACATCTACAGCAGGCTGTCAAGGCTGCT 82
QY 615 TAAGTCACTCGGCACCTCATCTGACAGTGAAGGACGCTCGGAGGCTGCCCATGCCCG 674
DB 81 TAAGTCACTCGGCACCTCATCTGACAGTGAAGGACGCTCGGAGGCTGCCCATGCCCG 22
QY 675 CACCACCTGTG 684
DB 21 CACCATTGTG 12

RESULT 10
BF953475 324 bp mRNA linear EST 22-JAN-2001
LOCUS RC3-NN1182-121100-011-902 NN1182 Homo sapiens cDNA, mRNA sequence.
DEFINITION
ACCESSION BF953475
VERSION BF953475.1 GI:12370750
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 324)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,

TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
10737800
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?ti=RC3&t3=RC3-NN1182-
121100-011-908&t3=2000-11-12&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 324.

FEATURES
source
1..324
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="NN1182"
/note="Organ: nervous normal; Vector: puc18; Site: 1: Smal;
Site 2: Smal; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

ORIGIN
Query Match 18.2%; Score 305.4; DB 10; Length 324;
Best Local Similarity 98.1%; Pred. No. 2.3e-49;
Matches 309; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 365 GTGACCCGAGGAGCACCTCTCTCAAGAGGGGATGAGAAAAGTGAACCTGG 424
DB 10 GTCACAACTTGAGCACCTCTCTCAAGAGGGGATGAGAAAAGTGAACCTGG 69
QY 425 TGCTGGGGAGCGCGGCTCTGGGCTCAGATCCGTGGGGAGCTGAGTACGGCTTG 484
DB 70 TGCTGGGGAGCGCGGCTCTGGGCTCAGATCCGTGGGGAGCTGAGTACGGCTTG 129
QY 485 GCATTTACATCACTGGGCTGGACCCAGGCTCTGAAGCAGAAGGACGCGGCTCAAGTTG 544
DB 130 GCATTTACATCACTGGGCTGGACCCAGGCTCTGAAGCAGAAGGACGCGGCTCAAGTTG 189
QY 545 GGGACCATTTCTAGAAGTGAATGGGGAGGCTTTCTCAACATCTCTACAGCAGGCTG 604
DB 190 GGGACCATTTCTAGAAGTGAATGGGGAGGCTTTCTCAACATCTCTACAGCAGGCTG 249
QY 605 TCAGGCTCTTAAGTCACTCTCGGCACCTCATCTCTGACAGTGAAGGACGCTGGGAGGCTGC 664
DB 250 TCAGGCTCTTAAGTCACTCTCGGCACCTCATCTCTGACAGTGAAGGACGCTGGGAGGCTGC 309
QY 665 CCCATGCCCGCACCA 679
DB 310 CCCATGCCCGCACCA 324

RESULT 11
BF953545 342 bp mRNA linear EST 22-JAN-2001
LOCUS RC3-NN1182-121100-012-904 NN1182 Homo sapiens cDNA, mRNA sequence.
DEFINITION

```

ACCESSION   BF953479
VERSION     BF953479.1
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens (human)

REFERENCE   1 (bases 1 to 342)
AUTHORS    Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
            O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
TITLE      Shotgun sequencing of the human transcriptome with ORF expressed
           sequence tags
JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE    20202663
PUBMED     10737800
COMMENT    Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC3&t2=RC3-NN1182-
            121100-012-h04&t3=2000-11-12&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 17
            High quality sequence stop: 342.
            Location/Qualifiers
              1..342
               /organism="Homo sapiens"
               /mol_type="mRNA"
               /db_xref="taxon:9606"
               /dev_stage="Adult"
               /clone_lib="NN1182"
               /note="Organ: nervous normal; Vector: puc18; Site 1: SmaI;
               Site 2: SmaI; A mini-library was made by cloning products
               derived from ORESTES PCR (U.S. Letters Patent application
               No. 196,716 - Ludwig Institute for Cancer Research)
               profiles into the pUC 18 vector. Reverse transcription of
               tissue mRNA and cDNA amplification were performed under
               low stringency conditions."

FEATURES             source
     source
     Query Match      18.0%; Score 301.8; DB 10; Length 342;
     Best Local Similarity 97.8%; Pred. No. 1.2e-48;
     Matches 306; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY      365  GTGACCGGAGGACCCCTGACCTCTGCAAGAGGGGATGAGAAAGGTGACCTGG 424
DB      30  GTCAACAATTGAGCACCCTGCACCTCTGCAAGAGGGGATGAGAAAGGTGACCTGG 89

QY      425  TGCTGGGGGACGGCGGCTCCCTGGGCTCAGATCCGTTGGGGAGCTGAGTACGGCCTTG 484
DB      90  TGCTGGGGGACGGCGGCTCCCTGGGCTCAGATCCGTTGGGGAGCTGAGTACGGCCTTG 149

QY      485  GCATTTACATCATCTGGCGTGAGACCCAGGCTCTGAAGCAGAGCAGCGGGCTCAAGGTTG 544
DB      150  GCATTTACATCATCTGGCGTGAGACCCAGGATCTGAAGCAGAGCAGCGGGCTCAAGGTTG 209

QY      545  GGGACCAAGATTCTAGAAGTGAATGGGGGGAGCTTTCTCAACATCTCTACAGCAGAGGCTG 604
DB      210  GGGACCAAGATTCTAGAAGTGAATGGGGGGAGCTTTCTCAACATCTCTACAGCAGAGGCTG 269

QY      605  TCAGGCTGCTTAAGTCACTCTGGCACTCTGAGTCAAGTGAAGGAGCTGGGAGGCTGC 664
DB      270  TCAGGCTGCTTAAGTCACTCTGGCACTCTGAGTCAAGTGAAGGAGCTGGGAGGCTGC 329

QY      665  CCCATGCCCGCAC 677
DB      330  CCCATGCCCGCAC 342

RESULT 12
LOCUS     BF953479
DEFINITION RC3-NN1182-121100-011-h06 NN1182 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF953479
VERSION   BF953479.1
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 343)
AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
           Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
           Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
           Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
           O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
           Simpson,A.J.
TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
          sequence tags
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE   20202663
PUBMED    10737800
COMMENT   Contact: Simpson A.J.G.
          Laboratory of Cancer Genetics
          Ludwig Institute for Cancer Research
          Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
          Brazil
          Tel: +55-11-2704922
          Fax: +55-11-2707001
          Email: asimpson@ludwig.org.br
          This sequence was derived from the FAPESP/LICR Human Cancer Genome
          Project. This entry can be seen in the following URL
          (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC3&t2=RC3-NN1182-
          121100-011-h06&t3=2000-11-12&t4=1)
          Seq primer: puc 18 forward
          High quality sequence start: 8
          High quality sequence stop: 343.
          Location/Qualifiers
            1..343
             /organism="Homo sapiens"
             /mol_type="mRNA"
             /db_xref="taxon:9606"
             /dev_stage="Adult"
             /clone_lib="NN1182"
             /note="Organ: nervous normal; Vector: puc18; Site 1: SmaI;
             Site 2: SmaI; A mini-library was made by cloning products
             derived from ORESTES PCR (U.S. Letters Patent application
             No. 196,716 - Ludwig Institute for Cancer Research)
             profiles into the pUC 18 vector. Reverse transcription of
             tissue mRNA and cDNA amplification were performed under
             low stringency conditions."

ORIGIN
Query Match      17.8%; Score 298.6; DB 10; Length 343;
Best Local Similarity 98.7%; Pred. No. 4.9e-48;
Matches 301; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      375  GAGCACCTTGACCTCTGCAAGAGGGGATGAGAAAGGTGACCTGGCTGGGGGA 434
DB      39  GAGCACCTTGACCTCTGCAAGAGGGGATGAGAAAGGTGACCTGGCTGGGGGA 98

QY      435  CGGCGGCTCCCTGGGCTCAGATCCGTTGGGGAGCTGAGTACGGCCTTGCATTACAT 494
DB      99  CGGCGGCTCCCTGGGCTCAGATCCGTTGGGGAGCTGAGTACGGCCTTGCATTACAT 158

QY      495  CACTGGCTGGACCCAGCTCTGAAGCAGAGGAGCGGGCTCAAGGTTGGGGACCAGAT 554

```

159 CACTGGCGTGACCCAGGCTCTGAAGCAGAGGAGCGGGCTCAAGGTTGGGACCAAGAT 218
555 TCTAGAAGTGAATGGCGGAGCTTTCTCAACATCCTACACGAGGAGCTGTCAAGGCTGCT 614
219 TCTAGAAGTGAATGGCGGAGCTTTCTCAACATCCTACACGAGGAGCTGTCAAGGCTGCT 278
615 TAAGTCATCTCGGCACCTCATCTGACAGTGAAGGAGCTCGGAGGCTGCCCATGCCCG 674
279 TAAGTCATCTCGGCACCTCATCTGACAGTGAAGGAGCTCGGAGGCTGCCCATGCCCG 338
675 CACCA 679
339 CACCA 343
RESULT 13
BF953480/c 308 bp mRNA linear EST 22-JAN-2001
LOCUS RC3-NN1182-121100-011-h10 NN1182 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF953480
ACCESSION BF953480.1 GI:12370755
VERSION EST.
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 308)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
10737800
PUBMED
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC3&t2=RC3-NN1182-
121100-011-h10&t3=2000-11-12&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 308.
Location/Qualifiers
1..308
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="NN1182"
/notes="Organ: nervous normal; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

17.5%; Score 293; DB 10; Length 308;
Best Local Similarity 99.7%; Pred. No. 6.1e-47;
Matches 304; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

ORIGIN

375 GAGCACCTTGACCTCTGCAAGGAGGGATGAGAAAAGGTGACCTGGTCTGGGGA 434
304 GAGCACCTTGACCTCTGCAAGGAGGGATGAGAAAAGGTGACCTGGTCTGGGGA 245
435 CGGCCGGTCCCTGGGCTCACGATCCGTGGGGAGCTGAGTACGGCCTTGGCATTTACAT 494
244 CGGCCGGTCCCTGGGCTCACGATCCGTGGGGAGCTGAGTACGGCCTTGGCATTTACAT 185
495 CACTGGCGTGAGCCAGGCTCTGAAGCAGAGGAGCGGCTCAAGGTTGGGAGCCAGAT 554
184 CACTGGCGTGAGCCAGGCTCTGAAGCAGAGGAGCGGCTCAAGGTTGGGAGCCAGAT 125
555 TCTAGAAGTGAATGGCGGAGCTTTCTCAACATCCTACACGAGGAGCTGTCAAGGCTGCT 614
124 TCTAGAAGTGAATGGCGGAGCTTTCTCAACATCCTACACGAGGAGCTGTCAAGGCTGCT 65
615 TAAGTCATCTCGGCACCTCATCTGACAGTGAAGGAGCTCGGAGGCTGCCCATGCCCG 674
64 TAAGTCATCTCGGCACCTCATCTGACAGTGAAGGAGCTCGGAGGCTGCCCATGCCCG 6
675 CACCA 679
5 CACCA 1
RESULT 14
BF953472
LOCUS RC3-NN1182-121100-011-d05 NN1182 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF953472
ACCESSION BF953472.1 GI:12370747
VERSION EST.
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 342)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
10737800
PUBMED
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC3&t2=RC3-NN1182-
121100-011-d05&t3=2000-11-12&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 342.
High quality sequence stop: 342.
Location/Qualifiers
1..342
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="NN1182"
/notes="Organ: nervous normal; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

17.5%; Score 293; DB 10; Length 308;
Best Local Similarity 99.7%; Pred. No. 6.1e-47;
Matches 304; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

ORIGIN

No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

ORIGIN

Query Match 17.5%; Score 293; DB 10; Length 342;
Best Local Similarity 99.7%; Pred. No. 6.1e-47;
Matches 304; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 375 GAGCACCTCGACCTCCCTGCAAGAGGGGATGAGAAAAGGTGAACCTGTGCTGGGGGA 434
Db 39 GAGCACCTCGCA-CTCCTGCAAGAGGGGATGAGAAAAGGTGAACCTGTGCTGGGGGA 97
QY 435 CGGCGGTCCCTGGGCTCAGATCCGTGGGGAGCTGAGTACGGCTTTGGCATTTACAT 494
Db 98 CGGCGGTCCCTGGGCTCAGATCCGTGGGGAGCTGAGTACGGCTTTGGCATTTACAT 157
QY 495 CACTGGCGTGGACCCAGGCTCTGAAGCAGAAAGGACGGGCTCAAGTTGGGGACCCAGAT 554
Db 158 CACTGGCGTGGACCCAGGCTCTGAAGCAGAAAGGACGGGCTCAAGTTGGGGACCCAGAT 217
QY 555 TTAGAGTGAATGGGCGGAGCTTTCTCAACATCTCAGCAGAGGCTGTCAAGCTGCT 614
Db 218 TTAGAGTGAATGGGCGGAGCTTTCTCAACATCTCAGCAGAGGCTGTCAAGCTGCT 277
QY 615 TTAGTTCATCTCGCACCTCATCTGACAGTGAAGACCTCGGAGGCTGCCCATGCCG 674
Db 278 TTAGTTCATCTCGCACCTCATCTGACAGTGAAGACCTCGGAGGCTGCCCATGCCG 337
QY 675 CACCA 679
Db 338 CACCA 342

RESULT 15

BF953542 323 bp mRNA linear EST 22-JAN-2001
LOCUS RC3-NN1182-121100-012-f04 NN1182 Homo sapiens cDNA, mRNA sequence.

DEFINITION BF953542

ACCESSION BF953542.1 GI:12370817

VERSION EST.

KEYWORDS Homo sapiens (human)

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 323)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, P.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V.,
O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

10737800

CONTACT: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the PAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC3&tl2=RC3-NN1182-121100-012-f04&tl3=2000-11-12&tl4=1)

Seq primer: puc 18 forward

High quality sequence stop: 323.

Location/Qualifiers

FEATURES

source

1..323
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="NN1182"
/note="Organ: nervous normal; Vector: puc18; Site: 1; Sma1;
Site 2: Sma1; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

ORIGIN

Query Match 17.4%; Score 292; DB 10; Length 323;
Best Local Similarity 100.0%; Pred. No. 9.6e-47;
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 388 CTCTCAAGAGGGGATGAGAAAAGGTGAACCTGTGCTGGGGACGGCGGTCCCTG 447
Db 32 CTCTCAAGAGGGGATGAGAAAAGGTGAACCTGTGCTGGGGACGGCGGTCCCTG 91
QY 448 GGCTTCACGATCCGTGGGGAGCTGAGTACGGGCTTTGACATTCATCCTGGCGTGGAC 507
Db 92 GGCTTCACGATCCGTGGGGAGCTGAGTACGGGCTTTGACATTCATCCTGGCGTGGAC 151
QY 508 CCAGGCTCTGAAGCAGAGGACGGGCTCAAGTTGGGGACGAGATTCTAGAGTGAAT 567
Db 152 CCAGGCTCTGAAGCAGAGGACGGGCTCAAGTTGGGGACGAGATTCTAGAGTGAAT 211
QY 568 GGGCGGAGCTTTCTCAACATCTTACACGACGAGGCTGTGAGGCTGCTTAAGTCAATCTCGG 627
Db 212 GGGCGGAGCTTTCTCAACATCTTACACGACGAGGCTGTGAGGCTGCTTAAGTCAATCTCGG 271
QY 628 CACCTCATCTGACAGTGAAGGACGTCGGAGGCTGCCCATGCCCGCACCA 679
Db 272 CACCTCATCTGACAGTGAAGGACGTCGGAGGCTGCCCATGCCCGCACCA 323

Search completed: April 22, 2004, 01:43:32

Job time : 4129.23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 21, 2004, 19:26:07 ; Search time 259.231 Seconds

(without alignments)
10930.600 Million cell updates/sec

Title: US-10-078-090-47

Perfect score: 667
Sequence: 1 gcgttggtcgccgaggt.....aagcgagaccaggggggaga 667

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_29Jan04: *
1: Geneseqn1980s: *
2: Geneseqn1990s: *
3: Geneseqn2000s: *
4: Geneseqn2001as: *
5: Geneseqn2001bs: *
6: Geneseqn2002s: *
7: Geneseqn2003as: *
8: Geneseqn2003bs: *
9: Geneseqn2003cs: *
10: Geneseqn2004s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	667	100.0	667	6	Abv83604 Human bre
2	647.2	97.0	1677	6	Abv83605 Human bre
3	76.8	11.5	710	6	Abt09598 Phase-1 R
4	76.8	11.5	3316	4	Ab110742 Drosophil
5	76.6	11.5	1817	5	Abal5256 Human ner
6	75.6	11.3	271990	4	Add25213 Fertility
7	73.6	11.0	2278	4	Ab112518 Drosophil
8	73.4	11.0	667	9	Add3798 Seabass p
9	73.4	11.0	67832	8	Ada02801 Mouse itk
10	73.4	11.0	67832	9	Adb72539 Mouse itk
11	73.4	11.0	67832	9	Adc85281 Mouse itk
12	73.2	11.0	742	6	Abv76748 Frog embr
13	73.2	11.0	42998	7	Ada14747 Human rib
14	73.2	11.0	42999	6	Abv65032 Invertebr
15	73.2	11.0	42999	9	Ada61411 Human rib
16	72.6	10.9	338	9	Add19844 Oreochrom
17	72.4	10.9	596	9	Add19952 Oreochrom
18	72.4	10.9	814	9	Add20177 Oreochrom
19	72.4	10.9	41907	8	Ada37416 Human rDN
20	72.2	10.8	534	9	Add20617 Oreochrom
21	72.2	10.8	110000	7	Continuation (6 of
22	72	10.8	448	2	Atc68682 Penaeus v
23	72	10.8	605	9	Add20225 Oreochrom

ALIGNMENTS

RESULT 1

ABV83604
ID ABV83604 standard; cDNA; 667 BP.

XX AC ABV83604;

XX DT 06-DEC-2002 (first entry)

XX DE Human breast specific gene SEQ ID NO 47.

XX KW Human; breast; cytostatic; cancer; transgenic; gene therapy; vaccine;

XX KW Gene; ss.

XX OS Homo sapiens.

XX PN WC200266605-A2.

XX PD 29-AUG-2002.

XX PF 14-FEB-2002; 2002WO-US004284.

XX PR 15-FEB-2001; 2001US-0268999P.

XX (DIAD-) DIADEXUS INC.

XX PI Salceda S, Macina RA, Hu P, Recipon H, Karra K, Cafferkey R;

XX PI Sun Y, Liu C;

XX DR WPI; 2002-713345/77.

XX PT New isolated breast specific nucleic acid molecules and polypeptides, useful for identifying, diagnosing, monitoring, staging, imaging and treating breast cancer and non-cancerous disease states in breast tissue.

XX PS Claim 1; Page 170-171; 254pp; English.

XX CC The invention relates to human breast specific nucleic acids (I)

XX CC comprising: (a) a sequence encoding any one of 95 protein sequences

XX CC (AB566614-ABP66708); (b) any one of 115 polynucleotide sequences

XX CC (ABV83558-ABV93672); (c) a molecule that selectively hybridizes to (a) or

XX CC (b); (d) a molecule having at least 60% sequence identity to (a) or (b).

XX CC The breast specific nucleic acid molecules, polypeptides and antibodies

XX CC are useful for identifying, diagnosing, monitoring, staging, imaging and

XX CC treating breast cancer and non-cancerous disease states in breast tissue.

XX CC They are also useful for producing transgenic animals and cells and

XX CC producing engineered breast tissue for treatment and research. The

Add19778 Gadus mor
Aai12688 Probe #26
Aba54378 Human foe
Aai34039 Probe #27
Aba43925 Human bre
Aba24157 Probe #26
Aak28111 Human bon
Aak02673 Human bra
Abs27717 Human liv
Aai02597 Probe #25
Abs02604 Human Gen
Abq2801 Tritium
Aai19220 Probe #91
Aba64228 Human foe
Aai44380 Probe #13
Aba31373 Probe #98
Aak38426 Human bon
Aak12706 Human bra
Abs38019 Human liv
Abs1484 Human Gen
Aah2391 Drosophil
Ada02579 Mouse Rel

C 24 71.8 10.8 299 9 ADD19778
25 71.8 10.8 475 4 Aai12688
26 71.8 10.8 475 4 ABA54378
27 71.8 10.8 475 4 Aai34039
28 71.8 10.8 475 4 ABA43925
29 71.8 10.8 475 4 ABA24157
30 71.8 10.8 475 4 AAK28111
31 71.8 10.8 475 4 AAK02673
32 71.8 10.8 475 4 ABS27717
33 71.8 10.8 475 5 AAI02597
34 71.8 10.8 475 6 ABS02604
35 71.8 10.8 527 6 ABQ2801
36 71.8 10.8 546 4 AAI19220
37 71.8 10.8 546 4 ABA64228
38 71.8 10.8 546 4 AAI44380
39 71.8 10.8 546 4 ABA31373
40 71.8 10.8 546 4 AAK38426
41 71.8 10.8 546 4 AAK12706
42 71.8 10.8 546 4 ABS38019
43 71.8 10.8 546 6 ABS1484
44 71.6 10.7 316 4 AAH2391
45 71.6 10.7 48995 8 ADA02579

RESULT 5
ABAI5256
ID ABAI5256 standard; DNA; 1817 BP.
XX
AC ABAI5256;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 7587.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisticking; antinaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antilucer; anticonvulsant; antifungal;
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001334.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 28-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225366P.
PR 14-AUG-2000; 2000US-0225367P.
PR 14-AUG-2000; 2000US-0225368P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 23-AUG-2000; 2000US-0227182P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232197P.
PR 14-SEP-2000; 2000US-0232198P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-023400P.
PR 14-SEP-2000; 2000US-023401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 23-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 20-OCT-2000; 2000US-0242221P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.

[illegible][illegible]

Qy 1 GGGTGGTCGCGGCCGAGGTCCATAACCTCGCCTCATCCAGATCTGTCAGATGAAG 60
| | | | | | | | | | | | | | | | | | | | | |
Dd 1 GGGTGGTCGCGGCCGAGGTCCATAACCTCGCCTCATCCAGATCTGTCAGATGAAG 60
| | | | | | | | | | | | | | | | | | | | | |
Qy 61 AGAGGGGAGGAGGGAAGAGAGATGCTTTGGGTGATTGTCAGAGGCCACCAG 120
| | | | | | | | | | | | | | | | | | | | | |

Db 61 AGAGGAGGAGAGGAGAGAGAGATGCTTTGGGGTGTATTTGGCCAGAGGCCACAG 120
 QY 121 GCTGATCCCATGAAGAAATCTGGTGAGAGGGTCTTAAAGTCATAAATGAGATCCAGT 180
 Db 121 GCTGATCCCATGAAGAAATCTGGTGAGAGGGTCTTAAAGTCATAAATGAGATCCAGT 180
 QY 181 TGCCAGGTGGTGCATAGTGCACACAGTGTAAATGTGTCACTTTTGTATCTTATCAGAA 240
 Db 181 TGCCAGGTGGTGCATAGTGCACACAGTGTAAATGTGTCACTTTTGTATCTTATCAGAA 240
 QY 241 ATCTCAGCTGGTGCCACCTGGCCAAATACACTGCAGAGCATGTCTGTCTGTCTGTG 300
 Db 241 ATCTCAGCTGGTGCCACCTGGCCAAATACACTGCAGAGCATGTCTGTCTGTCTGTG 300
 QY 301 TCTGTGTCTCTG 360
 Db 301 TCTGTGTCTCTG 360
 QY 361 TGTCTCTCATCTTTTCTATCTTATCATATAGTAGTATTAATAATAATTAGAGAGA 420
 Db 361 TGTCTCTCATCTTTTCTATCTTATCATATAGTAGTATTAATAATAATTAGAGAGA 420
 QY 421 TACACAGAAAATATAGAGAGATAACAGTGTCTCTTATAAAAAAAGAGCTGCCCT 480
 Db 421 TACACAGAAAATATAGAGAGATAACAGTGTCTCTTATAAAAAAAGAGCTGCCCT 480
 QY 481 CTCTCATAGCTTTTAACTACAGAACTCTTCGAGAAAAGAGACAAAACGGAGAAA 540
 Db 481 CTCTCATAGCTTTTAACTACAGAACTCTTCGAGAAAAGAGACAAAACGGAGAAA 540
 QY 541 CAAGAACAAACGGAGAGACAGTACAGAAACAGAGACAGCGGACAAAACACAGTGA 600
 Db 541 CAAGAACAAACGGAGAGACAGTACAGAAACAGAGACAGCGGACAAAACACAGTGA 600
 QY 601 GGGAGCAACACAGAGGGCGGACCAATTTACCCACACAGTGTGAAAAAGGAGACCCAGG 660
 Db 601 GGGAGCAACACAGAGGGCGGACCAATTTACCCACACAGTGTGAAAAAGGAGACCCAGG 660
 QY 661 GGGGAGA 667
 Db 661 GGGGAGA 667

RESULT 2

US-10-078-090-48
 ; Sequence 48, Application US/10078090
 ; Publication No. US20030044815A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Salceda, Susana
 ; APPLICANT: Macina, Roberto
 ; APPLICANT: Hu, Ping
 ; APPLICANT: Recipon, Herve
 ; APPLICANT: Karka, Kalpana
 ; APPLICANT: Cafferkey, Robert
 ; APPLICANT: Sun, Yongming
 ; APPLICANT: Liu, Chenghua
 ; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and P
 ; FILE REFERENCE: DEX-0312
 ; CURRENT APPLICATION NUMBER: US/10/078,090
 ; PRIOR FILING DATE: 2002-02-14
 ; PRIOR APPLICATION NUMBER: 60/268,999
 ; PRIOR FILING DATE: 2001-02-15
 ; NUMBER OF SEQ ID NOS: 210
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 48
 ; LENGTH: 1677
 ; TYPE: DNA
 ; ORGANISM: Homo sapien
 US-10-078-090-48

Query Match 97.0%; Score 647.2; DB 15; Length 1677;
 Best Local Similarity 99.5%; Pred. No. 4.1e-171;
 Matches 649; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 16 GAGGTCCATAACCCCTGCCCTCATCCAGATCTGTGCAGATGAAAGAGAGGGAGGAGAGG 75
 Db 1026 GGGGACCATTAACCTTGCCTCATCCAGATCTGTGCAGATGAAAGAGAGGGAGGAGAGG 1085
 QY 76 GAAAGAGAGAGATGCTTTGGGGTGTATTTGGCCAGAGGCCACCGCTGGATCCCATGAA 135
 Db 1086 GAAAGAGAGAGATGCTTTGGGGTGTATTTGGCCAGAGGCCACCGCTGGATCCCATGAA 1145
 QY 136 GAAATCTGGTGCAGAGGGTCTTAAAGTCTATAAATCTAGATCCAGTTGCGAGGTGCTGCA 195
 Db 1146 GAAATCTGGTGCAGAGGGTCTTAAAGTCTATAAATCTAGATCCAGTTGCGAGGTGCTGCA 1205
 QY 196 TAGTTGCCAACAGTGTATGTCTACCTTTTGTATCTTATCAGAAATCTCAGCCTGGTGG 255
 Db 1206 TAGTTGCCAACAGTGTATGTCTACCTTTTGTATCTTATCAGAAATCTCAGCCTGGTGG 1265
 QY 256 CCACCTGGCCAAATACACTGCAGAGCATGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 315
 Db 1266 CCACCTGGCCAAATACACTGCAGAGCATGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 1325
 QY 316 GTCTT 375
 Db 1326 GTCTT 1385
 QY 376 TCATCTTATCATTAACATAGTAGTATTAATAATAATTAGAGAGATACACAGAAAATATA 435
 Db 1386 TCATCTTATCATTAACATAGTAGTATTAATAATAATTAGAGAGATACACAGAAAATATA 1445
 QY 436 TAGAGAAGATAACAGTGTCTTCTATATAAAAAAAGAGCTGCCCTCTCTGTATAGTTCT 495
 Db 1446 TAGAGAAGATAACAGTGTCTTCTATATAAAAAAAGAGCTGCCCTCTCTGTATAGTTCT 1505
 QY 496 AACAATCTCAGTACTCTCGCAGAAAAGAGACAAAACGGGAGAAACAAAGAAACAAACGGG 555
 Db 1506 AACAATCTCAGTACTCTCGCAGAAAAGAGACAAAACGGGAGAAACAAAGAAACAAACGGG 1565
 QY 556 AGACAAGACTAGAGAAAACACAGGACAGCGGACAAAACCACTGAGGGAGCAACACACAGA 615
 Db 1566 AGACAAGACTAGAGAAAACACAGGACAGCGGACAAAACCACTGAGGGAGCAACACACAGA 1625
 QY 616 GGGGCGAACCATTAACCCACACAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 667
 Db 1626 GGGGCGAACCATTAACCCACACAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1677

RESULT 3

US-10-027-632-285026
 ; Sequence 285026, Application US/10027632
 ; Publication No. US20020198371A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, David G.
 ; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
 ; POLYMORPHISMS IN THE HUMAN GENOME
 ; FILE REFERENCE: 108827.129
 ; CURRENT APPLICATION NUMBER: US/10/027,632
 ; PRIOR FILING DATE: 2002-04-30
 ; PRIOR APPLICATION NUMBER: US 60/218,006
 ; PRIOR FILING DATE: 2000-07-12
 ; PRIOR APPLICATION NUMBER: US 60/198,676
 ; PRIOR FILING DATE: 2000-04-20
 ; PRIOR APPLICATION NUMBER: US 60/193,483
 ; PRIOR FILING DATE: 2000-03-29
 ; PRIOR APPLICATION NUMBER: US 60/185,218
 ; PRIOR FILING DATE: 2000-02-24
 ; PRIOR APPLICATION NUMBER: US 60/167,363
 ; PRIOR FILING DATE: 1999-11-23
 ; PRIOR APPLICATION NUMBER: US 60/156,358
 ; PRIOR FILING DATE: 1999-09-28
 ; PRIOR APPLICATION NUMBER: US 60/146,002
 ; PRIOR FILING DATE: 1999-08-09
 ; NUMBER OF SEQ ID NOS: 325720
 ; SOFTWARE: FastSeq for Windows Version 4.0

Search completed: April 22, 2004, 01:59:50
Job time : 287.003 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 21, 2004, 21:05:04 ; Search time 1636.77 Seconds

(without alignments)
12169.150 Million cell updates/sec

Title: US-10-078-090-47

Perfect score: 667

Sequence: 1 gcgtggtcgccgagcgggt.....aagcgagaccaggggggaga 667

Scoring table:

IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estnu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pin:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_red:*

26: em_gss_pig:*

27: em_gss_vrt:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	ID	Description
1	284.2	42.6	432	9 AL049078	AL049078 DKFZp434B
2	88.6	13.3	1246	9 AL514829	AL514829 AL514829
3	87.8	13.2	627	28 AZ997081	AZ997081 2M0283C01
4	87	13.0	752	28 BZ271973	BZ271973 CH230-275

86.6	13.0	715	9	AL663389
84.6	12.7	492	28	AZ897298
84	12.6	628	29	BX151392
83.8	12.6	851	14	CA581263
82.4	12.4	737	29	BX146272
82	12.3	796	28	BH350473
81.8	12.3	583	29	CNS044QG
81.2	12.2	740	29	CNS02YRZ
80.6	12.1	564	28	BH383538
80.6	12.1	635	28	BH370475
80.4	12.1	602	29	CC775881
79.8	12.0	595	29	DR44N19T
79.4	11.9	989	29	CNS01G9Q
79.2	11.9	474	12	BW582064
79.2	11.9	801	29	CNS01VC9
79	11.8	512	29	BX242790
79	11.8	575	29	BX182349
78.8	11.8	704	28	AQ780204
78.6	11.8	594	29	DR3L1S
78.4	11.8	528	13	BQ480630
78.4	11.8	543	28	BZ179552
78.4	11.8	758	28	BZ233572
78.4	11.8	880	29	BX240451
78.4	11.8	1159	9	AL514869
78.2	11.7	534	9	AL514819
78.2	11.7	608	28	AZ973484
78.2	11.7	687	28	AZ335248
78	11.7	551	9	AI586009
78	11.7	581	29	DR36D12T
78	11.7	776	28	BZ105168
78	11.7	1353	13	BX404043
77.6	11.6	345	14	CB937318
77.6	11.6	522	14	CF777662
77.6	11.6	544	28	AZ736516
77.4	11.6	442	29	CNS02E4P
77.4	11.6	491	28	AZ290346
77.4	11.6	674	29	BX196307
77.4	11.6	805	28	CC073103
77.4	11.6	809	28	CC080479
77.2	11.6	552	29	CNS03W90
77.2	11.6	638	28	AQ628460

ALIGNMENTS

RESULT 1	AL049078	432 bp	mRNA	linear	EST 04-SEP-2003
LOCUS	DKFZp434B1519_r1	434	(synonym: htes3)	Homo sapiens	cDNA clone
DEFINITION	DKFZp434B1519, mRNA sequence.				
ACCESSION	AL049078				
VERSION	AL049078.1				
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Ottewaelder, B., Obermaier, B., Mewes, H.W., Gassenhuber, J. and Wiemann, S.				
TITLE	EST (Ottewaelder, et al.)				
JOURNAL	Unpublished (1999)				
COMMENT	Contact: MIPS				

MIPS
Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by MediGenomix (Martinsried/Germany) within the cDNA sequencing consortium of the German Genome Project. No 5' sequence available.
This clone (DKFZp434B1519) is available at the RZPD in Berlin.

RESULT 5	AL663389	715 bp	linear	EST 11-JAN-2002
LOCUS	AL663389	directional larval cDNA	library	Ciona intestinalis CDNA
DEFINITION	clone 008ZD03 3', mRNA sequence.			
ACCESSION	AL663389			
VERSION	AL663389.1	GI:18130296		
KEYWORDS	EST.			
SOURCE	Ciona intestinalis			
ORGANISM	Ciona intestinalis			
REFERENCE	Sukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona; Phlebobranchia; Cloniidae; Ciona.			
AUTHORS	1 (bases 1 to 715)			
TITLE	Ciona intestinalis directional larval cDNA library			
JOURNAL	Genoscope.			
COMMENT	Unpublished (2002)			
CONTACT	Genoscope			
Genoscope - Centre National de Sequencage				
BP 191 91006 EVRY cedex - France				
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr				
IMPORTANT: this sequence may contain errors. The Ciona intestinalis library from which the clone was isolated may be contaminated with cDNAs from bacteria or other Eukarya.				
Directional larval cDNA library originate from Dr.M.Branno, Stazione A.Dohrn, Naples, Italy, and was prepared in pBluescript2SK+.				
Location/Qualifiers				
1. 715				
/organism="Ciona intestinalis"				
/mol_type="mRNA"				
FEATURES				
source				

[illegible]

SOURCE ORGANISM	REFERENCE AUTHORS	TITLE	JOURNAL MEDLINE PUBLISHED	COMMENT
Anopheles gambiae (African malaria mosquito)	Anopheles gambiae Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.	1 (bases 1 to 635) Hong, Y.S., Hogan, J.R., Wang, X., Sarkar, A., Sim, C., Loftus, B.J., Rep, C., Huff, E.R., Carlile, J.L., Black, K., Zhang, H.-B., Gardner, M.J. and Collins, F.H.		
		Construction of a BAC library and generation of BAC end sequence-tagged connectors for genome sequencing of the African malaria mosquito Anopheles gambiae	Mol. Genet. Genomics 258 (6), 720-728 (2003)	
			22542063	
			12655398	
		Other GSSs: AG-ND-180N9.TR		
		Contact: Brendan J Loftus		
		Department of Eukaryotic Genomics		
		The Institute for Genomic Research		
		9712 Medical Center Dr., Rockville, MD 20850, USA		
		Tel: 301 838 0208		
		Fax: 301 838 3543		

FEATURES
source
1. 635
Location/Qualifiers
Seq primer: M13 For
Class: BAC ends.
partial digest.
University, College Station, Texas 77843-2123, USA using a HindIII
library was constructed at Texas A&M University BAC Center
the gut. The DNA is derived from mixed sexes of larvae. The BAC
DNA. All DNA was extracted from newly hatched first instar larvae
(TIGR). The BAC library was generated from A. gambiae PEST strain
F.H. Collins and sequenced by The Institute for Genomic Research
This clone is from an A. gambiae BAC library (ND-TAX) provided by

[illegible]

Db
156 TGATGAGTGTGTGTGTGTGTGTGTGTGTGTGT 114

RESULT 15	CC775881	602 bp	DNA	linear	GSS 27-JUN-2000
LOCUS	CC775881/c				
DEFINITION	CH240_130L13.TV CHORI-240 Bos taurus genomic clone CH240_130L13, genomic survey sequence.				
ACCESSION	CC775881				
VERSION	CC775881.1	GI:32328172			
KEYWORDS	GSS.				
SOURCE	Bos taurus				
ORGANISM	Bos taurus (cow)				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;				
	Bovidae; Bovinae; Bos.				
REFERENCE	1 (bases 1 to 602)				
AUTHORS	Larkin,D.M. Everts-van der Wind,A., Rebeiz,M., Schweitzer,P.,				

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: April 21, 2004, 21:20:47 ; Search time 367 Seconds
(without alignments)
3009.623 Million cell updates/sec

Title: US-10-078-090-151
Perfect score: 1326
Sequence: 1 LRRKAHEGLGFSIRGSEH.....TMANSAGSHSARSNLTQTPG 260

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 337863 seqs, 2124099041 residues
Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xl
-Q/cgn2_1/USPTO.spool_P/US10078090/runat_20042004_132809_16579/app_query.fasta_1.455
-DB=N Geneseq 29Jan04 -QFMT=fastp -SUFFIX=p2n.rng -MINMATCH=0.1 -LOPCL=0
-LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10078090 @CCN 1.1.708 @runat_20042004_132809_16579 -NCPU=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 29Jan04:.*
1: Geneseqn1980s:.*
2: Geneseqn1990s:.*
3: Geneseqn2000s:.*
4: Geneseqn2001as:.*
5: Geneseqn2001bs:.*
6: Geneseqn2002s:.*
7: Geneseqn2003as:.*
8: Geneseqn2003bs:.*
9: Geneseqn2003cs:.*
10: Geneseqn2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1326	100.0	1677	6 ABV83605	ABV83605 Human bre
2	954	71.9	3102	3 AAC75441	AAC75441 Human ORF
3	706	53.2	1718	7 ADA53214	ADA53214 Human cod
4	438	33.0	2822	7 ABX71178	ABX71178 Novel hum
5	324	24.4	1338	4 ABL29731	ABL29731 Drosophil
6	321.5	24.2	2259	7 ABX34636	ABX34636 Human mdd
7	318.5	24.0	2162	2 AAX40058	AAX40058 Colon can
8	318.5	24.0	2236	2 AAX40057	AAX40057 Colon can

9	318.5	24.0	2236	2 AAX39644	AAX39644 Renal can
10	318.5	24.0	2289	2 AAX40061	AAX40061 Colon can
11	318.5	24.0	2409	2 AAX40062	AAX40062 Colon can
12	312	23.5	837	7 ABZ75386	ABZ75386 SID poly
13	290	21.9	6497	4 ABL29730	ABL29730 Drosophil
14	280	21.1	1489	4 ABL10215	ABL10215 Drosophil
15	231	17.4	3046	8 ACC85071	ACC85071 Human MBC
16	231	17.4	4108	8 AAD57342	AAD57342 Human kin
17	228	17.3	3071	7 ACA56540	ACA56540 Human sig
18	229	17.3	3071	8 ACC85072	ACC85072 Human MBC
19	228.5	17.2	3100	8 AAV20419	AAV20419 Human dis
20	228.5	17.2	3100	8 AAX86725	AAX86725 DNA seque
21	228.5	17.2	3100	8 ACC85073	ACC85073 Human MBC
22	228.5	17.2	3196	2 AAV20420	AAV20420 Human dis
23	228.5	17.2	5147	5 ABV25006	ABV25006 Human pro
24	226.5	17.1	4801	9 ADD78275	ADD78275 Human CGD
25	226.5	17.1	4884	8 AAD57243	AAD57243 Human CGD
26	226.5	17.1	4933	7 ACD13338	ACD13338 Human DNA
27	226.5	17.1	5079	8 AAD57244	AAD57244 Human CGD
28	226.5	17.1	5139	7 ABV72514	ABV72514 Nucleotid
29	222	16.7	3995	6 ABK35489	ABK35489 Human end
30	222	16.7	3995	6 ABK35548	ABK35548 Gene DLG4
31	222	16.7	3995	6 ABV78023	ABV78023 Hypoxia-r
32	222	16.7	3995	8 ACC85074	ACC85074 Human MBC
33	217.5	16.4	2963	2 AAT93775	AAT93775 PSD-93 co
34	217	16.4	3671	4 ABL10214	ABL10214 Drosophil
35	214	16.1	1274	4 AAT73917	AAT73917 Human PDZ
36	210	15.8	993	3 AAA30453	AAA30453 Human CDN
37	209	15.8	1492	9 ADB58401	ADB58401 Toxicity-
38	209	15.8	1492	9 ADB52979	ADB52979 Primary r
39	203.5	15.3	3454	4 ABL03235	ABL03235 Drosophil
40	203	15.3	4235	4 AAH98661	AAH98661 Human pro
41	203	15.3	4235	4 ABA08948	ABA08948 Human sem
42	199.5	15.0	4607	6 ABAK6540	ABK6540 DNA encod
43	198	14.9	8640	6 ABS63443	ABS63443 Human CDN
44	198	14.9	8640	6 ABS63444	ABS63444 Human CDN
45	197.5	14.9	873	6 ABK46532	ABK46532 Human PDZ

ALIGNMENTS

RESULT 1
ABV83605
ID ABV83605 standard; cDNA; 1677 BP.
XX AC ABV83605;
XX DT 06-DEC-2002 (first entry)
XX DE Human breast specific gene SEQ ID NO 48.
XX KW Human; breast; cytostatic; cancer; transgenic; gene therapy; vaccine;
XX OS Homo sapiens.
XX PN WO200266605-A2.
XX PD 29-AUG-2002.
XX PF 14-FEB-2002; 2002WO-US004284.
XX PR 15-FEB-2001; 2001US-0268999P.
XX PA (DIAD-) DIADEXUS INC.
XX PI Salceda S, Macina RA, Hu P, Recipon H, Karra K, Cafferkey R;
XX PI Sun Y, Liu C;
XX DR WPI; 2002-713345/77.
XX PT New isolated breast specific nucleic acid molecules and polypeptides,
XX PT useful for identifying, diagnosing, monitoring, staging, imaging and

treating breast cancer and non-cancerous disease states in breast tissue.

Claim 1; Page 171-172; 254pp; English.

The invention relates to human breast specific nucleic acids (I) comprising: (a) a sequence encoding any one of 95 protein sequences (ASP66614-ASP66708); (b) any one of 115 polynucleotide sequences (ABV83558-ABV83672); (c) a molecule that selectively hybridizes to (a) or (b); (d) a molecule having at least 50% sequence identity to (a) or (b). The breast specific nucleic acid molecules, polypeptides and antibodies are useful for identifying, diagnosing, monitoring, staging, imaging and treating breast cancer and non-cancerous disease states in breast tissue. They are also useful for producing transgenic animals and cells and producing engineered breast tissue for treatment and research. The transgenic animals are useful as animal model systems used in elaborating the biological function of the polypeptides, studying conditions and/or disorders associated with aberrant expression and in screening for compounds effective in ameliorating the conditions. The polynucleotides are useful for gene therapy and in vaccines.

Sequence 1677 BP: 407 A; 456 C; 473 G; 341 T; 0 U; 0 Other;

ment Scores:

Length:	6.24e-106	1677
Matches:	1326.00	260
Conservative:	100.00%	0
Mismatches:	100.00%	0
Indels:	100.00%	0
Gaps:	6	0

0-078-090-151 (1-260) x ABV83605 (1-1677)

1 LeuArgArgAlaIyAlaHisGluGlyLeuGlyPheSerIleArgGlyGlySerGluHis 20

4 TTGGGGGTGTCACAGGGCCACAGAGGCTTTGGGCTTCAGCATCCGTGGGGGCTCGAGAC 63

21 GlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLySerGluGlyLeu 40

64 GCGTGTGGGCATCTACGTGTCTCTGGTGAAACAGGCTCTTACGTGGAAGGAAGACTG 123

41 ArgValGlyAspGlnIleLeuArgValAsnAspLySerLeuAlaArgValThrHisAla 60

124 CGGGTCGGGGACCATTTCTGGCGTCAACGACAAATCCCTGGCCCGGGTGACCCACGCG 183

61 GluAlaValLyAlaLeuLySerLySerLySerLeuValLeuSerValTySerAlaGly 80

184 GAGGCGGTCAAGGCTCTGAAGGGCTCAGAAAGCTGGTGCTGTCTGTGTCTACAGAGG 243

81 ArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTyrValAspProGlnGlyArg 100

244 CCATCCCTGGGGGTACGTCAACCAACCATCTACACCTGGGTGGACCCCGCAGGGCGC 303

101 SerIleSerProSerGlyLeuProGlnProHisGlyAlaLeuArgGlnGlnGlu 120

304 AGCATCTCCCAACCCTCGGGGCTGCCCAAGCCAGCGTGGTGGCTCAGAGCAGCAGAG 363

121 GlyAspArgArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLyAlaValAsnLeu 140

364 GGTGACCGGAGAGACCCCTGCACCTCTGCAGAGGAGGGATGAGAAAAGGTGAACCTG 423

141 ValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGluTyrGlyLeu 160

424 GTGCTGGGGGACGGCCGGTCCCTGGGCTCACCGATCCGTGGGGAGCTGAGTACGGCTT 483

161 GlyIleTyrIleThrGlyValAspProGlySerGluAlaGluGlySerGlyLeuLyVal 180

484 GGCATTATCATCATCTGGCGGTGGACCCAGGCTCTGAAGCAGAAAGCAGCGGGCTCAAAGTT 543

181 GlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHisAspGluAla 200

544 GGGGACCATCTTATGAGAGTGAATGGGGCGAGCTTTCTCAACATCTCTACAGCAGAGCT 603

201 ValArgLeuLeuLySerSerArgHisIleLeuThrValLyAspValGlyArgLeu 220

Db	604	GTGAGCTGCTT	TAGTCA	TCTCGC	CACTCAT	CTGACAGTGAAGGAGCTCGGGAGCGCTG	240
Qy	221	ProHisAlaArgThrThrValAspGluThrIysTrpIleAlaSerSerArgIleAspGlu	240				
Db	664	CCCCATGCCCGCACCATGTGGACGAGACCAAGTGGATCCCACTGCCGATCCCGGATCAGGGAG	723				
Qy	241	ThrMetAlaAsnSerAlaGlySerGlyHisSerAlaArgSerAsnLeuGlnThrProGly	260				
Db	724	ACATGGCGAACTCGGCAGGGTCTGGCCACTCTGCTCGCTCCAATCTCCAGACCCAGGG	783				
RESULT 2							
AACT75441							
ID	AACT75441	standard; cDNA; 3102 BP.					
XX	XX	AACT75441;					
XX	XX	08-FEB-2001 (first entry)					
DT	DE	Human ORFX ORF996 polynucleotide sequence SEQ ID NO:1991.					
XX	XX	Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;					
KW	KW	vulnerable; antipsoriatic; antiparkinsonian; nontropic; neuroprotective;					
KW	KW	anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;					
KW	KW	immunosuppressant; thrombolytic; coagulant; vasotropic; antidiabetic;					
KW	KW	hypotensive; dermatological; immunosuppressive; antiinflammatory;					
KW	KW	antiviral; antibacterial; antifungal; antineumatic; antitubercid;					
KW	KW	antianemic; gene therapy; cancer; proliferative disorder; hypertension;					
KW	KW	neurodegenerative disorder; osteoarthritis; graft vs host disease;					
KW	KW	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;					
KW	KW	cholesterol ester storage; systemic lupus erythematosus; infection;					
KW	KW	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;					
KW	KW	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;					
KW	KW	bone damage; cartilage damage; antiinflammatory disease; coagulation;					
KW	KW	thrombosis; contraceptive; ss.					
XX	XX	Homc sapiens.					
OS	XX	WO200058473-A2.					
PN	XX	05-OCT-2000.					
PD	XX	31-MAR-2000; 2000WO-US008621.					
XX	XX	31-MAR-1999; 99US-0127607P.					
PR	XX	02-APR-1999; 99US-0127636P.					
PR	XX	05-APR-1999; 99US-0127728P.					
PR	XX	30-MAR-2000; 2000US-00540763.					
XX	XX	(CURA-) CURAGEN CORP.					
PA	XX	Shimkets RA, Leach M;					
PI	XX	WPI; 2000-602362/57.					
DR	XX	P-PSDB; ABA41232.					
DR	XX	Novel nucleic acids and peptides derived from open reading frame X,					
XX	XX	useful for treating e.g. cancers, proliferative disorders,					
PT	XX	neurodegenerative disorders and cardiovascular disease.					
PT	XX	Claim 5: Page 1517-1519; 5507op; English.					

XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease.
XX
XX
PS Claim 5; Page 1517-1519; 5307pp; English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
CC anticoagulant; antiarrhythmic; immunosuppressant; immunostimulant;
CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC antiviral; antifungal; antineoplastic; antitubercular; antihypertensive. The
CC sequences can be used for determining the presence of or predisposition
CC to, or preventing or treating pathological conditions associated with an
CC ORFX-associated disorder. The nucleic acids can be used to express ORFX

ropic;
; s.

```
    techika I;
```

atory
seases in
mpy.

	AGCAC	113
	lyLeu	40
	GACTG	119
	isAla	60

Db 1196 CGGGTCGGGGACCAATTCTGGCGCTCAACGACAAATCCCTGGCCCGGGTACCCACGCG 1255
 Qy 61 GluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAlaGly 80
 Db 1256 GAGGCGCTCAAGGCTCTGAAGGCTTCAAGAGCTGGTGGCTGTGTGTACTACAGAGG 1315
 Qy 81 ArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTyrValAspProGlnGlyArg 100
 Db 1316 CCATCTCTGGGGCTACGTACCAACACACATCTACACTGGGTGGAGCCCGCAGGGCGC 1375
 Qy 101 SerIleSerProProSerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlnGlu 120
 Db 1376 AGCACTCTCCCACTCGGGCTGGCCAGCCCAAGCTGGTGGCTGGCTGAGCAGCAGAG 1435
 Qy 121 GlyAspArgSerThrLeuHisLeuLeuGlnGlyGlyAspGlyLysValAsnLeu 140
 Db 1436 GTGACCGGAGGAGCACCTGCACTCTCTGCAAGAGGGGATGAGAAAAGTGAGTGGG 1495
 Qy 141 ValLeuGlyAspGlyArgSerLeu 148
 Db 1496 GTGGGAAAGGAGCGCCCTCTC 1519

RESULT 4

ABX71178

ID ABX71178 standard; cDNA; 2822 BP.

XX AC ABX71178;

XX DT 05-NAR-2003 (first entry)

XX DE Novel human cDNA sequence #403.

XX KW Human; Gene; ss; nervous system disorder; peripheral neuropathy;

XX KW Huntington's disease; amyotrophic lateral sclerosis; haemophilia;

XX KW neurodegenerative disease; Parkinson's disease; Alzheimer's disease;

XX KW autoimmune disease; systemic lupus erythematosus; rheumatoid arthritis;

XX KW insulin-dependent diabetes mellitus; anaemia; thrombocytopaenia; wound;

XX KW ulcer; burn; bone disorder; osteoporosis; osteoarthritis; stroke;

XX KW fibrosis; reperfusion injury; infection; allergic rhinitis; asthma;

XX KW Crohn's disease; anaphylaxis; proliferation; chemotactic;

XX KW differentiation; stem cell growth factor; haematopoiesis; chemokinetic;

XX KW haemostatic; antiinflammatory; expressed sequence tag; EST.

XX OS Homo sapiens.

XX PN WO200281731-A2.

XX PD 17-OCT-2002.

XX PF 29-JAN-2002; 2002WO-US0001222.

XX PR 30-JAN-2001; 2001US-00774528.

XX PA (HYSE-) HYSEQ INC.

XX PA (GOOD/) GOODRICH R. W.

XX PI Tang TY, Liu C, Zhou P, Asundi V, Zhang J, Zhao Q, Ren F;

XX PI Xue AJ, Yang Y, Wehrman T, Wang J, Wang D, Drmanac RT;

XX DR WPI; 2003-058563/05.

XX PT Novel polypeptide useful for treating neurodegenerative diseases, myeloid

XX PT or lymphoid cell disorders, bone disorders, mechanical and traumatic

XX PT disorders, coagulation disorders, and inflammatory diseases.

XX PS Claim 1; Page; 612pp; English.

XX CC This invention relates to the cDNA sequences encoding an isolated novel

XX CC human polypeptide. The protein encoded by the nucleic acid of the

XX CC invention is useful for treating central and peripheral nervous system

XX CC diseases (e.g. peripheral neuropathy, Huntington's disease, amyotrophic

CC lateral sclerosis); neurodegenerative diseases (e.g. Parkinson's disease,
 CC Alzheimer's disease); autoimmune disease (e.g. systemic lupus
 CC erythematosus, rheumatoid arthritis, insulin-dependent diabetes mellitus)
 CC ; myeloid or lymphoid cell disorders (e.g. anaemia and thrombocytopaenia)
 CC ; wounds, ulcers, burns; bone disorders (e.g. osteoporosis, osteoarthritis);
 CC mechanical and traumatic disorders (e.g. stroke, head trauma); lung or liver fibrosis;
 CC reperfusion injury in various tissues; bacterial, viral or fungal infections;
 CC allergic rhinitis; asthma; coagulation disorders (e.g. haemophilia);
 CC cancer and tumours; and inflammatory diseases (e.g. septic shock, Crohn's
 CC disease anaphylaxis). The protein may be used to inhibit the growth,
 CC infection or function of infectious agents such as bacteria, fungi,
 CC viruses, or to effect bodily characteristics, biorhythms or circadian
 CC cycles of rhythms. The protein may also have
 CC proliferation/differentiation, stem cell growth factor, haematopoiesis
 CC regulation, immune stimulating or suppressing, chemotactic/chemokinetic,
 CC haemostatic and thrombolytic, receptor/ligand, and antiinflammatory
 CC activities. The cDNA sequences of the invention are useful for expressing
 CC recombinant protein for analysis. The present sequence represents a novel
 CC human cDNA sequence of the invention, this sequence is an expressed
 CC sequence tag (EST) and was identified using subtractive hybridisation
 XX
 SQ Sequence 2822 BP; 557 A; 893 C; 922 G; 450 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 2,26e-28 Length: 2822
 Score: 438.00 Matches: 107
 Percent Similarity: 55.13% Conservative: 38
 Best Local Similarity: 40.68% Mismatches: 90
 Query Match: 33.03% Indels: 28
 DB: Gaps: 6

US-10-078-090-151 (1-260) x ABX71178 (1-2822)

Qy 1 LeuArgArgAlaLysAlaHisGluGly----LeuGlyPheSerIleArgGlySerGlu 19
 Db 324 GTCCGGGTGAGAGAGTCCAGCAGGAGGCTGGGCTTCAGCTGCGGGGCTCAGAG 383
 Qy 20 HisGlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGluGly 39
 Db 384 CATGGCTGGGCATCTCGTCAGCAAGTGGAGGAGGAGGAGTGCAGACGGGGTGGC 443
 Qy 40 LeuArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHis 59
 Db 444 CTGTGCGTGGGGGACCAAGATCACGGAGGTGAATGGGCTGAGCTGGAGACACCATG 503
 Qy 60 AlaGluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAla 79
 Db 504 GGTAGCGCCGTTAAGGTGCTGACACGACGAGCGCGCTGCACATGATGGTTCGGCGCATG 563
 Qy 80 GlyArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTrpValAspProGlnGly 99
 Db 564 GCGCGTGTGCGGGCATCAAGTTCTCCAGGAGAGACACACCGTGGGTGGATGTGTAAT 623
 Qy 100 ArgSerIle-----SerProProSerGlyLeuProGlnProHisGly 113
 Db 624 CGGCGCTGTAGTGGAGAGTGGGTTCAACACCTCCGAC----- 665
 Qy 114 GlyAlaLeuArgGlnGlnGluGlyAspArgArgSerThrLeuHisLeuLeuGlnGlyGly 133
 Db 666 -----ACCAGCTCAGAAAGATGGTGTCCGGCGC---ATCGTCCACCTATACACAACTCC 716
 Qy 134 AspGluLysLysValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArg 153
 Db 717 GACGAC-----TTCTGCTGGGCTTCAACATCCGT 746
 Qy 154 GlyGlyAlaGluTyrGlyLeuGlyIleTyrIleThrGlyValAspProGlySerGluAla 173
 Db 747 GGGGGCAAGGAGTTGGCCTGGGCATCTATGTGTCCAAAGTGGACCATGTGGTGGCTGCC 806
 Qy 174 GluGlySerGlyLeuLysValGlyAspGlnIleLeuValValAsnGlyArgSerPheLeu 193
 Db 807 GAGGAGAAATGGCATCAAGGTGGGGGACCAAGTCTCTGGACCAACGAGTGTGAGTTTGAC 866

XX PD 10-OCT-2002.
XX PF 27-MAR-2002; 2002WO-US009944.
XX PR 28-MAR-2001; 2001US-0279619P.
XX PR 29-MAR-2001; 2001US-0280067P.
XX PR 29-MAR-2001; 2001US-0280068P.
XX PR 16-MAY-2001; 2001US-0291280P.
XX PR 17-MAY-2001; 2001US-0291829P.
XX PR 17-MAY-2001; 2001US-0291849P.
XX PR 19-JUN-2001; 2001US-0299428P.
XX PR 20-JUN-2001; 2001US-0299776P.
XX PR 20-JUN-2001; 2001US-0300001P.
XX PA (INCY-) INCYTE GENOMICS INC.
XX PI Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Ansley SR;
PI Daugherty SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;
PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;
PI Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;
XX WPI: 2003-058431/05.
XX P-PADB; ABU11646.
XX PT New purified disease detection and treatment molecule proteins and
XX PT polynucleotides, useful for diagnosing, treating or preventing cancers
XX PT (e.g. leukemia or sarcoma), anemia, Crohn's disease, AIDS, osteoporosis
XX PT or hepatitis.
XX PS Claim 1; SEQ ID NO 197; 339pp + Sequence Listing; English.
XX CC This invention describes a novel disease detection and treatment molecule
XX CC polypeptide (MDPT) which has anti-inflammatory, immunosuppressive,
XX CC osteopathic, cytostatic, anti-HIV, haemostatic, nephrotropic,
XX CC antianemic, antipsoriatic and hepatotropic activity. The polynucleotides
XX CC and the polypeptides of the invention can be used for gene therapy.
XX CC protein replacement therapy and are useful for treating a variety of
XX CC diseases or conditions. These polypeptides or polynucleotides are
XX CC particularly useful for diagnosing, treating or preventing cell
XX CC proliferative disorders (e.g. cancers including adenocarcinoma,
XX CC leukemia, lymphoma, melanoma, myeloma or sarcoma), anaemia, Crohn's
XX CC disease, acquired immunodeficiency syndrome (AIDS), Goodpasture's
XX CC syndromes, inflammation, osteoporosis, thrombocytopenia, psoriasis or
XX CC hepatitis. ABX3440-ABX3483 encode the MDDT polypeptides represented in
XX CC ABU11450-ABU11845, described in the disclosure of the invention. NOTE:
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 2259 BP; 555 A; 598 C; 661 G; 445 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2,37e-18 Length: 2259
Score: 321.50 Matches: 94
Percent Similarity: 51.88% Conservative: 44
Best Local Similarity: 35.34% Mismatches: 102
Query Match: 24.25% Indels: 27
DB: 7 Gaps: 8

US-10-078-090-151 (1-260) x ABX34636 (1-2259)
QY 1 LeuArgAlaValAlaHis---GluGlyLeuGlyPheSerLeuArgGlyGlySerGlu 19
Db 355 GTGCGTCTGGACCGCTGCACCCGAGGCGCTGGGCTGAGTGTGGTGGCTGGAG 414
QY 20 HisGlyValGlyLeuValSerLeuValGluProGlySerLeuAlaGluGlyGlu 39
Db 415 TTGGGCTGGGCTCTCACTCCACCTCATCAAGGCGGCTCAGGACGACGAGCTCGG 474
QY 40 LeuArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHis 59

Db 475 CTCAGGTAGGGACGAGATCGTCGGATCAATGGATATTCATCTCTCTCTGATCCCAT 534
QY 60 AlaGluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAla 79
Db 535 GAGGAGTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 594
QY 80 GlyArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTrp-----ValAspPro 97
Db 595 GGCTGATCCCGTGAAGAGCTCTCTCTGATGAGCCCTCCTCTGGCAGTATGGAT--- 651
QY 98 GlnGlyArgSerIleSerProSerGlyLeuProGlnProHisGlyGlyAlaLeuArg 117
Db 652 -----CAGTTTGTGTGGAAATCTGGGGCGGTCGAGGCGAGCTGGGC----- 693
QY 118 GlnGlnGluGlyAspArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLys 137
Db 694 ---TCCCTCGAAATCGGAAAC-----AAGGAGAAGAAG 726
QY 138 ValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGlu 157
Db 727 GTCTTCATCAGCTGTAGGCTCCCGAGGCTTGGCTGCAGCATTTCCAGCGGCCCATC 786
QY 158 TyrGlyLeuGlyIleTyrIleThrGlyValAspProGlySerGluAlaGluGlySerGly 177
Db 787 CAGAGGCTTGGCATCTTTATCAGCCATGTGAAACCTGGCTCCCTGCTGCTGAGTGGGA 846
QY 178 LeuLysValGlyAspGlnIleLeuValAsnGlyArgSerPheLeuAsnIleLeuHis 197
Db 847 TTGAGATAGGGACGAGATTTGCGAAGTCAATGGCGTCGACTTCTTAACCTGGATCAC 906
QY 198 AspGluAlaValArgLeuLeuLysSerSerArgHisLeuIleLeuThr---ValLysAsp 216
Db 907 AAGGAGCTGTAAATGTGCTGAGAGTAGCGGAGCTGACCATCTCCATTGTGAGTGCA 966
QY 217 ValGlyArgLeuProHisAlaArgThrValAspGluThrLysTrpIleAlaSerSer 236
Db 967 GCTGGCGCG-----GAGCTGTTCATCAGACCGGCGGCGGTGGCAGAGCGG 1014
QY 237 ArgIleArgGluThrMetAlaAsnSerAlaGlySerGlyHisSerAlaArgSerAsnLeu 256
Db 1015 CGGAGGCTGAGCT-CCAGCGGCGAGGAGCTTCTCATGCAGACGCGCTGGCGATGGATC 1073
QY 257 Gln-----ThrProGly 260
Db 1074 CAACAGATCTCCAGGA 1091
RESULT 7
AA40058
ID AA40058 standard; DNA; 2162 BP.
XX AC AA40058;
XX DT 02-JUL-1999 (first entry)
XX DE Colon cancer associated gene.
XX KW Cancer associated antigen; diagnosis; research; treatment; human;
XX KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
XX KW prostate cancer; ss.
XX OS Homo sapiens.
XX PN WO9904265-A2.
XX PD 28-JAN-1999.
XX PF 15-JUL-1998; 98WO-US014679.
XX PR 17-JUL-1997; 97US-00896164.
XX PR 10-OCT-1997; 97US-0061599P.
XX PR 10-OCT-1997; 97US-0061765P.
XX PR 10-OCT-1997; 97US-00948705.
XX PR 11-OCT-1997; 97GB-00021697.

```

FR      22-JUN-1998;    98US--00102322.
XX
XX      (LUDW-) LUDWIG INST CANCER RES.
PA
XX
XX      Old LJ, Scanlan MJ, Stockert E, Gure A, Chen Y, Gout I;
PI      O'hare M, Obata Y, Pfeundschtuh M, Tureci O, Sahin U;
XX
XX      WPI; 1999-132448/11..
DR
XX
XX      New isolated cancer associated nucleic acids and polypeptides - isolated
PT      using sera from cancer patients, used to develop products for the
PPT     diagnosis, monitoring or treatment of cancers.
XX
XX      Claim 67; Page 658; 787pp; English.
XX
XX      The invention relates to a method for diagnosing a disorder characterised
CC      by expression of a human cancer associated antigen precursor coded for by
CC      a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC      biological sample isolated from a subject with an agent that specifically
CC      binds to the NAM, an expression product or a fragment of an expression
CC      product complexed with an HLA molecule; and (b) determining the
CC      interaction between the agent and the NAM or the expression product as a
CC      determination of the disorder. The products and methods can be used in
CC      the diagnosis, monitoring, research, or treatment of conditions
CC      characterised by the expression of various cancer associated antigens.
CC      The invention provides nucleic acid sequences and encoded polypeptides
CC      which are cancer associated antigen precursors expressed in human breast
CC      cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC      lung cancer
XX
XX      Sequence 2162 BP; 528 A; 573 C; 622 G; 439 T; 0 U; 0 Other;
SQ

Alignment Scores:
Pred. No.:          4, 1e-18          Length:          2162
Score:             318.50           Matches:         93
Percent Similarity: 51.88%           Conservative:   45
Best Local Similarity: 34.96%        Mismatches:    102
Query Match:       24.02%            Indels:        27
DB:                2                 Gaps:          8

US-10-078-090-151 (1-260) x AAX40058 (1-2162)
QY      1 LeuArgAlaValLeuAlaHis--GluGlyLeuGlyPhseSerIleArgGlyGlySerGlu 19
   :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      358 GTGCCTGTGGACCGTCTGCACCCGGAAGGCTGGCTGTGATGTGGTGCTGGCTGGAG 417
QY      20 HisGlyValGlyIleTyrValserLeuValGluProGlySerLeuAlaGlyLysGluGly 39
   :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      418 TTTCGCTGTGGCGCTCTTCATCTCCACTCATCAAGGCGGTGAGGACGACAGCGTCGGG 477
QY      40 LeuArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaargValThrHis 59
   :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      478 CTCACGGTAGGGGACGATCGTCGGATCAATGGATTCCCATCTCCCTCTGTACCCAT 537
QY      60 AlaGluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAla 79
   :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      538 GAGGAGGTCACTCAACTCATTCGAACCAAGAAAACCTGTGCCATCAAAGTGAGACACATC 597
QY      80 GlyArgIleProGlyGlyTy=ValThrAsnHisIleTyThrTrp-----ValAspPro 97
   :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      598 GGCTGTATCCCCGTGAAAGCTCTCTGATGAGCCCTCACTTGGCAGTATGNGAT--- 654
QY      98 GinglyArgSerIleSerProProSerGlyLeuProGlnProHsiGlyGlyAlaLeuArg 117
   :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      655 -----CAGTTTGTTCGGAATCTGGGGGCGTGGCGAGCGAGCGCTGGGC----- 696
QY      118 GingIngluGlyAspArgSerThrLeuHisLeuLeuGlnGlyGlyAspGlyLysLys 137
   :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      697 ----TCCCCTGGAAATCGGGAAAC-----AAGGAGAGAGAG 729
QY      138 ValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGlu 157
   :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      730 GTCTTCATCAGCTGTAGCTGCCGAGGCTTGCTGCAGCATTTCCAGCGCCCATC 789

```

QY	158	TyrGlyLeuGlyIleTyrIleThrcllyValAspProGlySerGluAlaGluGlySerGly	177
Db	790	CAGAAGCTGGCATCTTATCAGCCATGTGAACCTGGCTCCTCTCTCTGAGGTGGGA	849
QY	178	LeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHis	197
Db	850	TTGGAGATAGGGACACAGATTGTCAAGTCAATGGCGTCGACTTCTCTAAACCTGGATCAC	909
QY	198	AspGluAlaValArgLeuLeuLysSerArgHisLeuLeuThr--ValLysAsp	216
Db	910	AAGGAGGCTGTAAATGTGCTGAAATAAGCCGACCTGACCATCTCCATTGTAGCTGCA	969
QY	217	ValGlyArgLeuProHisAlaArgThrThrValAspGluThrLysTrpIleAlaSerSer	236
Db	970	GCTGGCGG-----CAGCTGTTTCATGACAGACCGGCGGCTGGCAGAGGCG	1017
QY	237	ArgIleArgGluThrMetAlaAsnSerAlaGlySerGlyHisSerAlaArgSerAsnLeu	256
Db	1018	CGCGAGCTGAGCT-GCAGCGCAGAGAGCTTCTATGCAAGAGCGGCTGGCGATGGAGTC	1076
QY	257	Gln-----ThrProGly	260
Db	1077	CAACAAGATCCTCCAGGA	1094
RESULT 8			
AX	AX40057	ID	AX40057 standard; DNA; 2236 BP.
XX	AX	AC	AX40057;
DT	02-JUL-1999		(first entry)
XX	XX		Colon cancer associated gene.
DE	XX		Cancer associated antigen; diagnosis; research; treatment; human;
KW	XX		breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW	XX		prostate cancer; ss.
XX	XX		Homo sapiens.
OS	XX		WO9904265-A2.
PN	XX		28-JAN-1999.
PD	XX		15-JUL-1998; 98WO-US014679.
PF	XX		17-JUL-1997; 97US-00896164.
XX	XX		10-OCT-1997; 97US-0061599P.
PR	XX		10-OCT-1997; 97US-0061765P.
PR	XX		10-OCT-1997; 97US-00948705.
PR	XX		11-OCT-1997; 97GB-00021697.
PR	XX		22-JUN-1998; 98US-00102322.
XX	XX		(LUDW-) LUDWIG INST CANCER RES.
XX	XX		Old LJ, Scanlan MJ, Stockert E, Gure A, Chen Y, Gout I;
PI	XX		O'hare M, Obata Y, Pfreundschuh M, Tureci O, Sahin U;
XX	XX		WPI; 1999-132448/11.
DR	XX		New isolated cancer associated nucleic acids and polypeptides - isolated
XX	XX		using sera from cancer patients, used to develop products for the
PT	XX		diagnosis, monitoring or treatment of cancers.
PT	XX		Claim 67; Page 655-656; 787pp; English.
PS	XX		The invention relates to a method for diagnosing a disorder characterised
XX	XX		by expression of a human cancer associated antigen precursor coded for by
CC	XX		a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC	XX		biological sample isolated from a subject with an agent that specifically
CC	XX		binds to the NAM, an expression product or a fragment of an expression
CC	XX		product complexed with an HLA molecule; and (b) determining the

QY 1 LeuArgAlaLysAlaHis---GluGlyLeuGlyPheSerIleArgGlySerGlu 19
 Db GTGGCTGTGGACCGCTGACCCGGAAGGCTCGGCTGAGTGGCTGGCTGGAG 417
 QY 20 HisGlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGlyGly 39
 Db TTTGGCTGTGGCTGTCTTCTCCCTCCCTCATCAAAAGGGGCTGAGCAGACAGCTCGGG 477
 QY 40 LeuArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHis 59
 Db CTCCAGGTAGGGACGAGATGCTCGGATCAATGGATATTCCATCTCTCTCTCTCTCTCT 537
 QY 60 AlaGluAlaValLysAlaLeuLysGlySerLysLeuValLeuSerValTyrSerAla 79
 Db GAGAGGTCTACCACTCTGCAACCAAGAAACTGTCTCCATCAAAAGTGGAGACATC 597
 QY 80 GlyArgIleProGlyGlyTyrValThrAsnHisIleTyrThr 97
 Db GGCCTGATCCCGTGAAGAGCTCTCTGATGAGCCCTCATCTGGCAGTATGGAT--- 654
 QY 98 GlnGlyArgSerIleSerProSerGlyLeuProGlnProHisGlyAlaLeuArg 117
 Db ---CAGTTTGTGCGGAATCTGGGGGCTGCGAGGAGCGCTGGC----- 696
 QY 118 GlnGlnGluGlyAspArgArgSerThrLeuHisLeuGlnGlyAspGlyLys 137
 Db ---TCCCTCGAATCGGAAGAAC-----AAGGAGAAGAAG 729
 QY 136 ValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyAlaGlu 157
 Db GTCTTCTACCTGCTGAGCTCCGAGGCTTGGCTGAGCATTTCCAGCGCCCTCATC 789
 QY 158 TyrGlyLeuGlyIleTyrIleThrGlyValAspProGlySerGluAlaGlySerGly 177
 Db CAGAGGCTGGCATCTTTATCAGCCATGTGAACCTGGCTCTCTCTCTCTCTCTCTCT 849
 QY 178 LeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHis 197
 Db TTGAGATAGGGACCAAGATTGCAAGTCAATGGGCTGACCTTCTTAACTGATCATC 909
 QY 198 AspGluAlaValArgLeuLeuLysSerArgHisLeuLeuThr---ValLysAsp 216
 Db AAGGAGGCTGTAAATGTCTGAAAAATAGCCGAGCTGACCATCTCCATTTAGCTGCA 969
 QY 217 ValGlyArgLeuProHisAlaArgThrValAspGluThrLysTrpIleAlaSerSer 236
 Db GTCTGGCGG-----GAGCTTCTATGACAGACCGGAGGCTGCGAGGGG 1017
 QY 237 ArgIleArgGluThrMetAlaAsnSerAlaGlySerGlyHisSerAlaArgSerLeu 256
 Db CGGAGGCTGAGCT-GCAGCGGAGGAGCTTCTCATGCAAGCGGCTGGCGATGAGTGC 1076
 QY 257 Gln-----ThrProGly 260
 Db CAACAAGATCTCCAGGA 1094
 RESULT 10
 AAX40061
 ID AAX40061 standard; DNA; 2289 BP.
 XX
 AC AAX40061;
 XX
 DT 02-JUL-1999 (first entry)
 XX
 DE Colon cancer associated gene.
 XX
 KW Cancer associated antigen; diagnosis; research; treatment; human;
 KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
 KW prostate cancer; ss.
 XX
 OS Homo sapiens.
 XX
 FN W09904265-A2.

XX 28-JAN-1999.
 XX 15-JUL-1998; 98WO-US014679.
 XX 17-JUL-1997; 97US-00896164.
 PR 10-OCT-1997; 97US-0061599P.
 PR 10-OCT-1997; 97US-0061765P.
 PR 10-OCT-1997; 97US-0094870S.
 PR 11-OCT-1997; 97GB-00021697.
 PR 22-JUN-1998; 98US-00102322.
 XX (LUDW-) LUDWIG INST CANCER RES.
 PA Old Lz, Scanlan MJ, Stockert E, Gure A, Chen Y, Gout I;
 FI O'hare M, Obata Y, Pfreundschuh M, Tureci O, Sahin U;
 XX WPI; 1999-132448/11.
 XX New isolated cancer associated nucleic acids and polypeptides - isolated
 PT using sera from cancer patients, used to develop products for the
 PT diagnosis, monitoring or treatment of cancers.
 XX Claim 67; Page 663-664; 787pp; English.
 PS The invention relates to a method for diagnosing a disorder characterised
 CC by expression of a human cancer associated antigen precursor coded for by
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
 CC biological sample isolated from a subject with an agent that specifically
 CC binds to the NAM, an expression product or a fragment of an expression
 CC product complexed with an HLA molecule; and (b) determining the
 CC interaction between the agent and the NAM or the expression product as a
 CC determination of the disorder. The products and methods can be used in
 CC the diagnosis, monitoring, research, or treatment of conditions
 CC characterised by the expression of various cancer associated antigens.
 CC The invention provides nucleic acid sequences and encoded polypeptides
 CC which are cancer associated antigen precursors expressed in human breast
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
 CC lung cancer.
 XX
 SQ Sequence 2289 BP; 567 A; 596 C; 668 G; 458 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 4,4e-18 Length: 2289
 Score: 318.50 Matches: 93
 Percent Similarity: 51.88% Conservative: 45
 Best Local Similarity: 34.96% Mismatches: 102
 Query Match: 24.02% Indels: 27
 DB: 2 Gaps: 8
 US-10-078-090-151 (1-260) x AAX40061 (1-2289)
 QY 1 LeuArgAlaLysAlaHis---GluGlyLeuGlyPheSerIleArgGlySerGlu 19
 Db GTGGCTGTGGACCGCTGACCCGGAAGGCTCGGCTGAGTGGCTGGCTGGAG 417
 QY 20 HisGlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGlyGly 39
 Db TTTGGCTGTGGCTGTCTTCTCCCTCCCTCATCAAAAGGGGCTGAGCAGACAGCTCGGG 477
 QY 40 LeuArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHis 59
 Db CTCCAGGTAGGGACGAGATGCTCGGATCAATGGATATTCCATCTCTCTCTCTCTCTCT 537
 QY 60 AlaGluAlaValLysAlaLeuLysGlySerLysLeuValLeuSerValTyrSerAla 79
 Db GAGAGGTCTACCACTCTGCAACCAAGAAACTGTCTCCATCAAAAGTGGAGACATC 597
 QY 80 GlyArgIleProGlyGlyTyrValThrAsnHisIleTyrThr 97
 Db GGCCTGATCCCGTGAAGAGCTCTCTGATGAGCCCTCATCTGGCAGTATGGAT--- 654
 QY 98 GlnGlyArgSerIleSerProSerGlyLeuProGlnProHisGlyAlaLeuArg 117

DB 655 -----CAGTTTGTGCGAATCTGGGCGCTGGCAGGCGAGCTGGGC----- 696
 QY 118 GlnGlnGlnGlyAspArgSerThrLeuHisLeuGlnGlyGlyAspGluLys 137
 DB 697 ---TCCCTGGAAATCGGAAAC-----AAGGAGAGAG 729
 QY 138 ValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyAlaGlu 157
 DB 730 GTCTTCATCAGCTGTAGCTCCGAGGCTTGGCTGAGCATTTCCAGGCGGCCCATC 789
 QY 158 TyrGlyLeuGlyIleThrGlyValAspProGlySerGluAlaGluGlySerGly 177
 DB 790 CAGAGCCTGGCATCTTTATCAGCCATGTGAACCTGGCTCCCTGCTGAGTGGA 849
 QY 178 LeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHis 197
 DB 850 TTGGAGATAGGAGCAGATTGTGAGTCAATGGCTGACCTTCTTAACCTGATCAC 909
 QY 198 AspGluAlaValArgLeuLeuLysSerSerArgHisLeuLeuThr---ValLysAsp 216
 DB 910 AAGGAGGCTGTAATGTGCTGAAATAGCCGAGCTGACCATCTCCATTTGATGCTCA 969
 QY 217 ValGlyArgLeuProHisAlaArgThrThrValAspGluThrLysTrpIleAlaSerSer 236
 DB 970 GCTGGCGG-----GAGCTGTTCATGACAGCCGGGAGCGGCTGCGAGGCG 1017
 QY 237 ArgIleArgGluThrMetAlaAsnSerAlaGlySerGlyHisSerAlaArgSerAsnLeu 256
 DB 1018 CGGACGCTGAGCT-CCAGCGGCGAGGCTTCTCATGACAGAGCGGCTGGCGATGAGTC 1076
 QY 257 Gln-----ThrProGly 260
 DB 1077 CAACAAGATCCTCCAGGA 1094

RESULT 11
 AAX40062
 ID AAX40062 standard; DNA; 2409 BP.
 XX AC AAX40062;
 XX DT 02-JUL-1999 (first entry)
 XX DE Colon cancer associated gene.
 XX KW Cancer associated antigen; diagnosis; research; treatment; human;
 XX KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
 XX KW prostate cancer; ss.
 XX OS Homo sapiens.
 XX PN WO9904265-A2.
 XX PD 28-JAN-1999.
 XX PF 15-JUL-1998; 98WO-US014579.
 XX PR 17-JUL-1997; 97US-00896164.
 XX PR 10-OCT-1997; 97US-0061599P.
 XX PR 10-OCT-1997; 97US-0061765P.
 XX PR 10-OCT-1997; 97US-00948705.
 XX PR 11-OCT-1997; 97GB-00021697.
 XX PR 22-JUN-1998; 98US-00102322.
 XX PA (LUDW-) LUDWIG INST CANCER RES.
 XX PI Old LJ, Scanlan MJ, Stockert E, Gure A, Chen Y, Gout I;
 XX PI O'hare M, Obata Y, Pfeundschn M, Tureci O, Sahin U;
 XX DR WPI; 1999-132448/11.
 XX PT New isolated cancer associated nucleic acids and polypeptides - isolated
 PT using sera from cancer patients, used to develop products for the

PT diagnosis, monitoring or treatment of cancers.
 XX Claim 67; Page 665-666; 787pp; English.
 XX The invention relates to a method for diagnosing a disorder characterised
 CC by expression of a human cancer associated antigen precursor coded for by
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
 CC biological sample isolated from a subject with an agent that specifically
 CC binds to the NAM, an expression product or a fragment of an expression
 CC product complexed with an HLA molecule; and (b) determining the
 CC interaction between the agent and the NAM or the expression product as a
 CC determination of the disorder. The products and methods can be used in
 CC the diagnosis, monitoring, research, or treatment of conditions
 CC characterised by the expression of various cancer associated antigens.
 CC The invention provides nucleic acid sequences and encoded polypeptides
 CC which are cancer associated antigen precursors expressed in human breast
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
 CC lung cancer
 XX Sequence 2409 BP; 595 A; 638 C; 692 G; 484 T; 0 U; 0 Other;
 SQ Alignment Scores: Length: 2409
 Pred. NO.: 4.68e-18 Matches: 93
 Score: 318.50 Conservative: 45
 Percent Similarity: 51.88% Mismatches: 102
 Best Local Similarity: 34.96% Indels: 27
 Query Match: 24.02% Gaps: 8
 DB: 2
 US-10-078-090-151 (1-260) x AAX40062 (1-2409)
 QY 1 LeuArgArgAlaLysAlaHis---GluGlyLeuGlyPheSerIleArgGlySerGlu 19
 DB 358 GTGGCTCTGGACGCTCTGCACCCGGAAGGCTCGGCTGAGTGTGGTGGCTGGAG 417
 QY 20 HisGlyValGlyIleThrValSerLeuValGluProGlySerLeuAlaGluLysGluGly 39
 DB 418 TTTGGCTGTGGGCTCTTCATCTCCACCTCATCAAGGCGGCTCAGGAGAGCGGTCGG 477
 QY 40 LeuArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHis 59
 DB 478 CTCAGGTAGGAGCAGATCGTCGGATCAATGATATTCATCTCTCTGTACCAT 537
 QY 60 AlaGluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAla 79
 DB 538 GAGAGGTCTCAACCTCATTCGAAACCAAGAACTGTGTCATCAAGTGAGACATC 597
 QY 80 GlyArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTrp-----ValAspPro 97
 DB 598 GGCCTGATCCCGTGAAGAGCTCTCTGATGAGCCCTCACTTGGCAGTATGTGAT--- 654
 QY 98 GlnGlyArgSerIleSerProProSerGlyLeuProGlnProHisGlyAlaLeuArg 117
 DB 655 -----CAGTTTGTGCGAATCTGGGCGCTGGGAGGAGCGCTGGGC----- 696
 QY 118 GlnGlnGlnGlyAspArgArgSerThrLeuHisLeuGlnGlyGlyAspGluLysLys 137
 DB 697 ---TCCCTGGAAATCGGAAAC-----AAGGAGAGAG 729
 QY 138 ValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyAlaGlu 157
 DB 730 GTCTTCATCAGCTGTAGGCTCCGAGGCTTGGCTGAGCATTTCCAGGCGGCCCATC 789
 QY 158 TyrGlyLeuGlyIleThrGlyValAspProGlySerGluAlaGluGlySerGly 177
 DB 790 CAGAGCCTGGCATCTTTATCAGCCATGTGAACCTGGCTCCCTGCTGAGTGGA 849
 QY 178 LeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHis 197
 DB 850 TTGGAGATAGGAGCAGATTGTGAGTCAATGGCTGACCTTCTTAACCTGATCAC 909
 QY 198 AspGluAlaValArgLeuLeuLysSerSerArgHisLeuLeuThr---ValLysAsp 216

Db 910 AAGGAGGCTGTAATGTGCTGAAATAAGCCAGCCGTGACCATCTCCATTGTAGCTCA 969
 Qy 217 ValGlyArgLeuProHisAlaArgThrThrValaspGluThrLysTrpIleAlaSerSer 236
 Db 970 GCTGCGCGG-----GAGCTGTTCATGACAGACCGGCGGCTGGCAGAGCG 1017
 Qy 237 ArgIleArgGluThrMetAlaAsnSerAlaGlySerGlyHisSerAlaArgSerAsnLeu 256
 Db 1018 CGGACGCTGAGCT-CCAGCGGCGAGGAGCTTCTCATGACAGAGCGGCTGGCGATGAGTC 1076
 Qy 257 Gln-----ThrProGly 260
 Db 1077 CAACAAGATCTCCAGGA 1094
 RESULT 12
 ID ABZ75586 standard; DNA; 837 BP.
 AC ABZ75586;
 XX
 XX 28-APR-2003 (first entry)
 DT SID polypeptide encoding DNA.
 DE
 DE
 XX
 KW SID; deafness; hearing disorder; auditory; gene therapy; human; ds.
 XX
 OS Unidentified.
 FH Key Location/Qualifiers
 FT 1. .837
 FT /*tag a
 FT /product= "SID"
 FT
 PN W02003000891-A2.
 XX
 XX 03-JAN-2003.
 XX
 XX 21-JUN-2002; 2002WO-EP008065.
 XX
 PR 21-JUN-2001; 2001US-0299848P.
 PR 05-FEB-2002; 2002EP-00290277.
 XX
 XX (HYBR-) HYBRIGENICS.
 PA (INSP) INST PASTEUR.
 XX
 XX Daviet L, Legrain P, Petit C, Boeada B, El-Amraoui A;
 XX WPI; 2003-201427/19.
 DR P-PSDB; ABP71383.
 XX
 PT New complexes of Selected Interacting Domains of polypeptides or
 PT polynucleotides, useful for preventing and treating deafness and other
 PT inner ear or hearing disorders in humans or animals.
 XX
 PS Claim 13; Page 62-63; 92pp; English.
 CC
 CC The invention relates to protein-protein interactions involved in
 CC deafness or hearing disorders, specifically proteins that interact with
 CC mammalian inner ear proteins. Complexes of selected interacting domains
 CC (SID) of polypeptides are useful in preventing and treating deafness and
 CC other inner ear or hearing disorders and/or diseases in humans or
 CC animals. The methods are useful in screening drugs for agents which
 CC modulate protein interaction. The present sequence represents a DNA
 CC encoding the SID polypeptide that interacts with the human myosin VIIa
 XX
 SQ Sequence 837 BP; 201 A; 211 C; 270 G; 155 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 4.74e-18 Length: 837
 Score: 312.00 Matches: 83
 Percent Similarity: 51.53% Conservative: 35
 Best Local Similarity: 36.24% Mismatches: 83
 Query Match: 23.53% Indels: 28

DB: 7 Gaps: 6
 US-10-078-090-151 (1-260) x ABZ75586 (1-837)
 Qy 8 GluGlyLeuGlyPheSerIleArgGlyGlySerGluHisGlyValGlyIleTyrValSer 27
 Db 19 GAAGGTCTCGGCTCAGCTGCGTGGAGCCTGGAAATTTGGCTGTGGATCTTTATCTCC 78
 Qy 28 LeuValGluProGlySerLeuAlaGluLysGluGlyLeuArgValGlyAspGlnIleLeu 47
 Db 79 CACCTCATCAAGGTGGCCAGGACAGAGCTTGGCTTCAGGTAGGGATGAATGTC 138
 Qy 48 ArgValAsnAspLysSerLeuAlaArgValThrHisAlaGluAlaValLysAlaLeuLys 67
 Db 139 CGGATCAACGGCTATTCCATCTCTCTGCTACCATGAGGAAGTCAACACCTGATCCGC 198
 Qy 68 GlySerLysLysLeuValLeuSerValTyrSerAlaGlyArgIleProGlyGlyTyrVal 87
 Db 199 ACCAAGNAGACCGTGTCCCAAGTGAGACACATCGGACTGATCCCTGTGAAGAGCTCT 258
 Qy 88 ThrAsnHisIleTyrThrTrp-----ValAspProGlnGlyArgSerIleSerProPro 105
 Db 259 CCGAGGAGTCCCTCAAAATGGCAGTATGTGGAT-----CAGTTGCTGCGGAATCT 309
 Qy 106 SerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlnGluGlyAspArgArgSer 125
 Db 310 GGGGTGTG-----CGAGTGGCTTG-----GGCTCACCTGGCAATCGGACAC 354
 Qy 126 ThrLeuHisLeuLeuGlnGlyGlyAspGluLysValAsnLeuValLeuGlyAspGly 145
 Db 355 -----NAGGAGAAGAAGGTGTTTATCAGTCTAGTGGGCTCT 390
 Qy 146 ArgSerLeuGlyLeuThrIleArgGlyGlyAlaGluTyrGlyLeuGlyIleTyrIleThr 165
 Db 391 CGGCGCTGGGCTGACGATCTCCAGTGGCCCATCCAGAAAGCTGGCATCTTCGTGAGC 450
 Qy 166 GlyValAspProGlySerGluAlaGluGlySerGlyLeuLysValGlyAspGlnIleLeu 185
 Db 451 CACGTGAAGCCTGGCTCCCTGCTCTGACAGAGTGGGGTTAGAGACAGAGACAGATTGTG 510
 Qy 186 GluValAsnGlyArgSerPheLeuAsnIleLeuHisAspGluAlaValArgLeuLys 205
 Db 511 GAAGTCAATGGCATAGACTTCACCACTGGACCAAGAGGCTGTGAATGTCTCTGAAG 570
 Qy 206 SerSerArgHisLeuIleLeuThrVal-----Lys 215
 Db 571 AGCAGCCGACGCTGACCATCTCCATCGTTGCGGCGCGGAGGCTGTTTCATGACG 630
 Qy 216 AspValGlyArgLeuProHisAlaArg 224
 Db 631 GACCGGGAACGGCTGGAGAGGACCG 657
 RESULT 13
 ID ABL29730/c
 XX ABL29730 standard; DNA; 6497 BP.
 AC ABL29730;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 40663.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 PD 27-SEP-2001.
 XX
 XX 23-MAR-2001; 2001WO-US009231.
 XX

QY 131 GlnGlyGlyAspGluLysValAsnLeuValLeu-----GlyAspGlyArgSerLeu 148
Db 574 -----GTTGGCAGCGGACATAGTGTGTCTCATGTGGCGCCAGCAAGAGCTG 627
QY 149 GlyLeuThrIleArgGlyGlyValAlaGluTyrGlyLeuGlyIleTyrIleThrGlyValAsp 168
Db 628 GGACTGGGATTTGCAAGGACCCGAGTGAAGCGGGGCATCTTCGTGCAGTTCCACAAAG 687
QY 169 ProGlySerGluAlaGluGlySerGlyLeuLysValGlyAspGlnIleLeuGluValAsn 188
Db 688 GATCGGAGTGTGCGCCGGAGCGGGATTACGGCCAGGTGACCAAGTCTCAGCGTCAAC 747
QY 189 GlyArgSerPheLeuAsnIleLeuHisAspGluAlaValArgLeuLeuLysSerSerArg 208
Db 748 AGCATCGACTTCTCGGACGTCTCTTTAGCGAGGCGCGTGGCGGTGATCAGAGCAGCAGC 807
QY 209 HisLeuIleLeuThrValLys 215
Db 808 AAGCTGGACATGGTGTGCGC 828
RESULT 15
ACC85071
ID ACC85071 standard; DNA; 3046 BP.
XX AC ACC85071;
XX DT 13-OCT-2003 (first entry)
XX DE Human MBCAT polypeptide encoding DNA.
XX KW MBCAT; beta-catenin; cytosolic; gene therapy; cancer; human; gene; ds.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX CDS 189..2969
XX FT /*tag= a
XX FT /product= "MBCAT"
XX PN WO2003052068-A2.
XX PD 26-JUN-2003.
XX PF 12-DEC-2002; 2002WO-US039796.
XX PR 13-DEC-2001; 2001US-0340213P.
XX PR 13-DEC-2001; 2001US-0340314P.
XX PR 13-DEC-2001; 2001US-0340322P.
XX PR 15-FEB-2002; 2002US-0357502P.
XX PA (EXEL-) EXELIXIS INC.
XX PI Costa MA, Gendreau SB, Dora EG, Nicoll M;
XX DR WPI; 2003-533010/50.
XX DR P-PSDB; ABR82220.
XX PT Identifying a candidate beta-catenin pathway modulating agent for
PT diagnosing or treating cancer by detecting a test agent-biased activity
PT of the assay system comprising a purified MBCAT polypeptide or nucleic
PT acid.
XX PS Example; Page 49-51; 81pp; English.
XX CC The invention relates to genes that modify beta-catenin pathway and to
CC the identification of human MBCAT (modifiers of beta-catenin) genes. The
CC MBCAT polypeptides are therapeutic targets for disorders associated with
CC beta-catenin function and are useful for manufacturing a medicament for
CC diagnosing or treating breast, colon, lung or ovary cancer. The present
CC sequence represents a human MBCAT polypeptide encoding DNA
XX Sequence 3046 BP; 975 A; 612 C; 681 G; 778 T; 0 U; 0 Other;

Alignment Scores: 2,578-10 Length: 3046
Pred. No.: 231.00 Matches: 91
Score: 46.91% Conservative: 38
Best Local Similarity: 33.09% Mismatches: 100
Query Match: 17.42% Indels: 46
DB: 8 Gaps: 15
US-10-078-090-151 (1-260) x ACC85071 (1-3046)
QY 1 LeuArgArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGly-----GlySer 18
Db 1146 ATAAGCTCATTAAGAGTCTTAAGGTCCTTGGGTTTAGCATTTGCTGGAGGTGTTGGAAT 1205
QY 19 GluHisGlyValGly-----IleTyrValSerLeuValGluProGlySerLeuAla 35
Db 1206 GAGCATATTCCTGGGATTAATAGCATCTAIGTACCAAAATTAATGAAGAGGTGCAGCA 1265
QY 36 GluLysGluGly---LeuArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeu 54
Db 1266 CATAGGATGGCAAACTTCAGATTGGAGATAAACTTTACAGAGTGAATAACGATATGTTTA 1325
QY 55 AlaArgValThrHisAlaGluAlaValLysAlaLeuLysGlySerLysLysLeuVal--- 73
Db 1326 GAAGAAGTTACTCATGAAGACGACTGCTTAAAGAACACATCTCAATTTGTTTAT 1385
QY 74 -----LeuSerValTyrSerAlaGlyArgIleProGlyGlyTyr--- 86
Db 1386 TTGAAGTGGCAAAACCCACAAAGTATGAT-----ATGAATGATGGCTATGCA 1433
QY 87 -----ValThrAsnHisIleTyrThrTpValAspProGlnGlyArgSerIleSer 103
Db 1434 CCACCTGATATCACCAAC-----TCTTCTTCAGCCTGTTGATAACCATGTTAGC 1484
QY 104 ProProSerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlnGluGlyAspArg 123
Db 1485 CCATCTTCCTTCTTGGCCAGACACCCAGCATCTCCAGCCAGATACTCCCCAGTTTCTAAA 1544
QY 124 ArgSerThrLeuHisLeuLeuGlnGlyGlyAspGlu-----LysLysVal 138
Db 1545 -----GCAGTACTTGGAGATGATGAATTTACAAGGGAACCTAGAAAAGTT 1589
QY 139 AsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGluTyr 158
Db 1590 GTTCTTCATCGTGGC---TCAAGGGCCCTTGGTTTCAACATTTGAGGAGA---GAAGAT 1643
QY 159 GlyLeuGlyIleTyrIleThrGlyValAspProGlySerGluAlaGluGlySerGly--- 177
Db 1644 GGAGAAAGAAATATTTATTTCCITTTATCTTAGCCGAGGACCTGCTGATCTAAGTGGAGAG 1703
QY 178 LeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHis 197
Db 1704 CTCAGAAAGGAGATCGTATTATATCGGTAAACAGTGTGTGACCTCAGAGCTGCTAGTCAT 1763
QY 198 AspGluAlaValArgLeuLeuLysSerSerArgHisLeuIleLeuThrValLysAspVal 217
Db 1764 GAGCAGGCAGCAGCTGCATTGAAAATGCTGGCCAGGTGTCACAATTTGTCACAATAT 1823
QY 218 GlyArgLeuProHisAlaArgThrThrValAspGluThrLysTrpIleAlaSerSerArg 237
Db 1824 CGACCTGAAGATACACATCGTTTT-----GAAGCTAAA-----ATACATGAT 1865
QY 238 IleArgGluThrMetAlaAsn-----SerAlaGlySerGly 249
Db 1866 TTACGGGAGCAGATGATGATAGTAGTAGTATTAGTTACAGGTCAGGT 1910
Search completed: April 22, 2004, 02:06:17
Job time : 381 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 22, 2004, 01:43:37 ; Search time 388 Seconds
(without alignments)
3021.245 Million cell updates/sec

Title: US-10-078-090-151

Perfect score: 1326

Sequence: 1 LRRKAHEGLGFSIRGSEH.....TVANSAGSGHARSNLOTTPG 260

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2907579 seqs, 2254313464 residues

Total number of hits satisfying chosen parameters: 5815158

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DRV=xlp
-Q=/cgn2_1/USPTO.spool.p/US10078090/runat_20042004_132811_16613/app.query.fasta_1.455
-DB=Published Applications NA -QMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS-human40.cdi -LIST=45 -DOCLALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORV=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10078090@cgn_1_133 @runat_20042004_132811_16613
-NCPU=6 -ICPU=3 -NO MAP -WARGQUERY -NEG SCORES=0 -WAIT -DSFBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:
1: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/prodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/prodata/1/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/prodata/1/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/prodata/1/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/prodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/prodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq:
9: /cgn2_6/prodata/1/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/prodata/1/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/prodata/1/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/prodata/1/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/prodata/1/pubpna/US09_NEW_PUB.seq:
14: /cgn2_6/prodata/1/pubpna/US10A_PUBCOMB.seq:
15: /cgn2_6/prodata/1/pubpna/US10B_PUBCOMB.seq:
16: /cgn2_6/prodata/1/pubpna/US10C_PUBCOMB.seq:
17: /cgn2_6/prodata/1/pubpna/US10_NEW_PUB.seq:
18: /cgn2_6/prodata/1/pubpna/US60_NEW_PUB.seq:
19: /cgn2_6/prodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

ALIGNMENTS

RESULT 1

US-10-078-090-48
; Sequence 48, Application US/10078090
; Publication No. US20030044815A1

GENERAL INFORMATION:

; APPLICANT: Saleda, Susana

; APPLICANT: Macina, Roberto

; APPLICANT: Hu, Ping

; APPLICANT: Recipon, Hervé

; APPLICANT: Karra, Kalpana

; APPLICANT: Cafferkey, Robert

; APPLICANT: Sun, Yongming

; APPLICANT: Liu, Chenghua

; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and P

; FILE REFERENCE: DEX-0312

; CURRENT APPLICATION NUMBER: US/10/078.090

; CURRENT FILING DATE: 2002-02-14

; PRIOR APPLICATION NUMBER: 60/268,999

; PRIOR FILING DATE: 2001-02-15

; NUMBER OF SEQ ID NOS: 210

; SOFTWARE: Patentin version 3.1

; SEQ ID NO 48

; LENGTH: 1677

1	1326	100.0	1677	15	US-10-078-090-48	Sequence 48, Appl
2	951.5	71.8	2964	13	US-10-112-944-536	Sequence 536, App
3	706	53.2	1718	16	US-10-094-749-782	Sequence 782, App
4	438	30.8	2822	16	US-10-120-988-406	Sequence 406, App
5	408.5	30.8	599	15	US-10-029-386-3728	Sequence 3728, App
6	346	26.1	195	15	US-10-029-386-17428	Sequence 17428, A
7	312	23.5	837	15	US-10-177-191A-5	Sequence 5, Appli
8	229	17.3	3071	16	US-10-305-720-1138	Sequence 1138, Ap
9	226.5	17.1	5139	13	US-10-451-207-14	Sequence 14, Appl
10	222	16.7	3995	9	US-09-919-497-9	Sequence 9, Appli
11	222	16.7	3995	13	US-10-170-385-300	Sequence 300, App
12	203	15.3	4235	13	US-10-275-774-724	Sequence 724, App
13	203	15.3	4235	13	US-10-296-115-486	Sequence 486, App
14	198	14.9	8640	13	US-09-964-956-24	Sequence 24, Appl
15	198	14.9	8640	13	US-09-964-956-26	Sequence 26, Appl
16	197.5	14.9	5836	10	US-09-998-425-2	Sequence 2, Appli
17	197.5	14.9	5836	10	US-09-997-977-2	Sequence 2, Appli
18	197	14.9	7516	9	US-09-917-800A-1664	Sequence 1664, Ap
19	196.5	14.8	2000	13	US-10-342-887-1009	Sequence 1009, Ap
20	196.5	14.8	2000	13	US-10-172-118-1009	Sequence 1009, Ap
21	196.5	14.8	2000	15	US-10-177-293-427	Sequence 427, App
22	196.5	14.8	2411	9	US-09-823-356-23	Sequence 23, Appl
23	196	14.8	1578	9	US-09-880-107-3440	Sequence 3440, Ap
24	195	14.7	1750	15	US-10-210-120-80	Sequence 80, Appl
25	195	14.7	3180	16	US-10-104-047-796	Sequence 796, App
26	195	14.7	6582	13	US-10-342-887-958	Sequence 958, App
27	195	14.7	6582	13	US-10-172-118-958	Sequence 958, App
28	195	14.7	7431	9	US-09-951-401-2	Sequence 2, Appli
29	195	14.7	7431	9	US-09-922-101-2	Sequence 2, Appli
30	195	14.7	7431	9	US-09-951-402-2	Sequence 2, Appli
31	194	14.6	3259	14	US-10-001-843-71	Sequence 71, Appl
32	191	14.4	6383	9	US-09-728-952-68	Sequence 68, Appl
33	189.5	14.3	1152	15	US-10-106-698-1545	Sequence 1545, Ap
34	189	14.3	3170	13	US-10-114-270-75	Sequence 75, Appl
35	186.5	14.1	1905	13	US-10-210-281-79	Sequence 79, Appl
36	186.5	14.1	3520	13	US-10-362-892-27	Sequence 27, Appl
37	186.5	14.1	3520	16	US-10-288-798-27	Sequence 27, Appl
38	186.5	14.1	5640	15	US-10-393-892-9	Sequence 9, Appli
39	186.5	14.1	5640	15	US-10-394-382-9	Sequence 9, Appli
40	186.5	14.1	6588	13	US-10-112-944-99	Sequence 1259, Ap
41	186.5	14.1	8062	13	US-10-342-887-1259	Sequence 1259, Ap
42	186.5	14.1	8062	13	US-10-172-118-1259	Sequence 45, Appl
43	186.5	14.1	8119	9	US-09-802-669-45	Sequence 45, Appl
44	186.5	14.1	8119	13	US-10-619-220-45	Sequence 3, Appli
45	183	13.8	3885	10	US-09-836-499-3	

```

; TYPE: DNA
; ORGANISM: Homo sapien
US-10-078-090-48

Alignment Scores:
Pred. No.:      2,748-145      Length:      1677
Score:          1326.00      Matches:      260
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      100.00%      Indels:      0
DB:              15          Gaps:      0

US-10-078-090-151 (1-260) x US-10-078-090-48 (1-1677)
QY 1 LeuAArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGlyGlySerGluHis 20
Db 4 TTGGCGGCGTCAAGGCGCCCAAGGCGCTTGGGCTTCAGCATCCGTGGGGCTCGGAGCAC 63
QY 21 GlyValGlyLeuValSerLeuValGluProGlySerLeuAlaGluGlySerLeu 40
Db 64 GCGGTGGGCGATCTACGTGTCTCTGGTGGAAACAGGCTCTCTAGCTGAGAAGAGGACTG 123
QY 41 ArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHis 60
Db 124 CGGGTGGGAGCACAGATTCTGGCGTCAAGACAAATCTCTGGCGGGTACCACGCG 183
QY 61 GluAlaValLysAlaLeuLysGlySerLysLeuValLeuSerValTyrSerLadly 80
Db 184 GAGCGCGTCAAGGCTCTGAAGGCTCCAAAGAGCTGGTCTGTGTGTACTCACAGGG 243
QY 81 ArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTyrValAspProGlnGlyArg 100
Db 244 CGCATCTCCCGGCGTACGTACACACACACATCTACACCTGGTGGTGAACCCGCGGCGCG 303
QY 101 SerIleSerProSerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlnGlu 120
Db 304 AGCATCTCCCGGCGTCCGCGCTCCAGGCGCCACGCGTGGTGGTGGTGGTGGTGGTGG 363
QY 121 GlyAspArgArgSerThrLeuHisLeuGlnGlyGlyAspGlyLysValAsnLeu 140
Db 364 GGTGACCGAGAGACACCTCTGACCTCTGCAAGGAGGGGATGAGAAAAGGTGAACCTG 423
QY 141 ValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGluTyrGlyLeu 160
Db 424 GTGCTGGGGGACGCGCGTCCCTGGGCTCCAGCATCCGTGGGGAGCTGAGTACGCGCTT 483
QY 161 GlyIleTyrIleThrGlyValAspProGlySerGluAlaGluGlySerGlyLeuLysVal 180
Db 484 GGCATTACATCTGCGTGGACCCAGGCTCTGAGCAGAGAGGCGGCGCTCAAGGTT 543
QY 181 GlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHisAspGluAla 200
Db 544 GGGGACGAGATTCTAGAAGTGAATGGCGGAGGCTTTCTCAACATCTCTACACGAGGCT 603
QY 201 ValArgLeuLeuLysSerSerArgHisLeuIleLeuThrValLysAspValGlyArgLeu 220
Db 604 GTCAGGTGCTTAAGTCACTCTGGGACCTCATCTGACATGAGTGAAGACGTCGGAGGCTG 663
QY 221 ProHisAlaArgThrThrValAspGluThrIleTyrPheIleAlaSerSerArgIleArgGlu 240
Db 664 CCCCATGCCGACCACTGTGGACGAGACCAAGTGGATCGCCAGTTCGCCGATCAGGGAG 723
QY 241 ThrMetAlaAsnSerAlaGlySerGlyHisSerAlaArgSerAsnLeuGlnThrProGly 260
Db 724 ACCATGGCGAATCGGCGAGGCTGTGGGACCTCTGCTCGCTCCATCTCCAGACCCCGAGG 783

RESULT 2
US-10-112-944-536
; Sequence 536 Application US/10112944
; Publication No. US20040048249A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Yang, Yonghong
; TITLE OF INVENTION: Secreted Polypeptides
; FILE REFERENCE: 805A
; CURRENT APPLICATION NUMBER: US/10/112,944
; CURRENT FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/515,126
; PRIOR FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: US 09/519,705
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: US 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: US 09/552,929
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 09/577,408
; PRIOR FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 924
; SOFTWARE: pt FL_genes Version 5.0
; SEQ ID NO 536
; LENGTH: 2964
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-112-944-536

Alignment Scores:
Pred. No.:      4,79e-101      Length:      2964
Score:          951.50      Matches:      191
Percent Similarity: 89.95%      Conservative: 6
Best Local Similarity: 87.21%      Mismatches: 15
Query Match:      71.76%      Indels:      7
DB:              13          Gaps:      2

US-10-078-090-151 (1-260) x US-10-112-944-536 (1-2964)
QY 43 GlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAlaGluAla 62
Db 20 GCGCCACGACTACTTCTCTCTGAGTGGCTGCTGTGTGTACTCAGTGCC-----TGT 61
QY 63 ValLysAlaLeuLysGlySerLysLeuValLeuSerValTyrSerAlaGlyArgIle 82
Db 62 GTCCAGGCTCTGAAGGCTCCCAAGAGCTGGTGTGTGTGTACTCAGCAGGCGGCATC 121
QY 83 ProGlyGlyTyrValThrAsnHisIleTyrThrTrpValAspProGlnGlyArgSerIle 102
Db 122 CCTGGGGGCTAGCTCACCACCATCTACACCTGGGTGGACCCGCGGCGGCGGAGCATC 181
QY 103 SerProProSerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlnGlyAsp 122
Db 182 TCCCCACCTCTGGGCTCTGCCCCAGCCCGTGGTGGTGGTGGTGGTGGTGGTGGTGG 241
QY 123 ArgArgSerThrLeuHisLeuLeuGlnGlyGlyAspGlyLysValAsnLeuValLeu 142
Db 242 CGGAGGAGCACCTCTCTCTGCAAGGAGGGGATGAGAAAAGGTGAACCTGTGTGTGT 301
QY 143 GlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGluTyrGlyLeuGlyIle 162
Db 302 GGGGACGCGCGCTCTCTCTGGGCTCAGCATCCGTGGGGGAGCTGAGTACGGCCTTGGCATT 361
```


QY 163 TyrIleThrGlyValAspProGlySerGluAlaGluGlySerGlyLeuLysValGlyAsp 182
DB 362 TACATCACTGCGTGGACCCAGGCTCTGAACACAGAGCGCGGCTCAAGTTGGGAC 421
QY 183 GinIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHisAspGluAlaValArg 202
DB 422 CAGATTCTAGAGTGAATGGCGGAGCTTTCTCAACATCTCTACACGACGAGGCTGTCAAG 481
QY 203 LeuLeuLysSerSerArgHisIleLeuThrValLysAspValGlyArgLeuProHis 222
DB 482 CTGCTTAAGTCACTCTCGGCACTCTATCTGACAGTGAAGACGTCGGGAGGCTGCCCAT 541
QY 223 AlaArgThrValAspGluThrLysTrpIleAlaSerSerArgIleArgGluThrMet 242
DB 542 GCCCGCACCACTGTGGACGAGACCAAGTGGATCGCCAGTTCCTCCGATCAGGAGACCATG 601
QY 243 AlaAsnSerAlaGly---SerGlyHisSerAlaArgSerAsnLeuGlnThrProGly 260
DB 602 GCGAACTCGGACAGGTTTCTTGGCGATCTCAACACAGAGGAATAACACAGCCAGGA 658

RESULT 3

US-10-094-749-782
; Sequence 782, Application US/10094749
; Publication No. US20030219741A1

GENERAL INFORMATION:

; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 782
; LENGTH: 1718
; TYPE: DNA
; ORGANISM: Homo sapiens

US-10-094-749-782

Alignment Scores:
Pred. No.: 1,388-72 Length: 1718
Score: 706.00 Matches: 140
Percent Similarity: 95.95% Conservative: 2
Best Local Similarity: 94.59% Mismatches: 6
Query Match: 53.24% Indels: 0
DB: 16 Gaps: 0

US-10-078-090-151 (1-260) x US-10-094-749-782 (1-1718)

QY 1 LeuArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGlyGlySerGluHis 20
DB 1076 TTCCGGCGTGCCCAAGCCACGAGGCGTTGGGCTTCAGCATCCGTGGGGCTCGGAGC 1135

QY 21 GlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGluGlyLeu 40
DB 1136 GCGTGGGCACTACGTGTCTCTGGTGAACCCAGGCTCTCTAGCTGAGAAGGAGGACTG 1195
QY 41 ArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAla 60
DB 1196 CCGGTTCGGGACCAACATCTCGCGCTCAACGACAAATCCCTGGCCCGGGTGACCCACGG 1255
QY 61 GluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAlaGly 80
DB 1256 GAGCGCGTCAAGGCTCTGAAGGGCTCCAAGAAAGTGGTGTCTGTGTGTACTCAGCAGG 1315
QY 81 ArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTrpValAspProGlnGlyArg 100
DB 1316 CGCATCTCTGGGGGCTTACCTCACCAACCACTTACACCTGGTGGACCCGCGGCGCC 1375
QY 101 SerIleSerProProSerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlnGlu 120
DB 1376 AGCATCTCCCAACCTCGGGCTCGCCACGCCACCGGTGGTGGCCCTGAGGACGAGGAG 1435
QY 141 ValLeuGlyAspGlyArgSerLeu 148
DB 1496 GTGGGGAAGGAGGCCAGCCTCTC 1519

RESULT 4

US-10-120-988-406
; Sequence 406, Application US/10120988
; Publication No. US20030219745A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Goodrich, Ryle
; APPLICANT: Liu, Chenghua
; APPLICANT: Ren, Feiyan
; APPLICANT: Wang, Dunrui
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20030219745A1el Nucleic Acids and
; FILE REFERENCE: 802CON
; CURRENT APPLICATION NUMBER: US/10/120,988
; CURRENT FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: 09/774,528
; PRIOR FILING DATE: 2001-01-30
; NUMBER OF SEQ ID NOS: 441
; SOFTWARE: Pf-FL_genes Version 2.0
; SEQ ID NO 406
; LENGTH: 2822
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (66)..(2822)
US-10-120-988-406

Alignment Scores:
Pred. No.: 7,15e-41 Length: 2822
Score: 438.00 Matches: 107
Percent Similarity: 55.13% Conservative: 38
Best Local Similarity: 40.68% Mismatches: 90
Query Match: 33.03% Indels: 28
DB: 16 Gaps: 6

US-10-078-090-151 (1-260) x US-10-120-988-406 (1-2822)

QY 1 LeuArgAlaLysAlaHisGluGly---LeuGlyPheSerIleArgGlyGlySerGlu 19
DB 324 GTCCGGGTGGAGAGACTCCAGCAGGAGGCTGGGCTTCAGCGTGGCGGGGCTCAGAG 383
QY 20 HisGlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGluGly 39

Db 384 CATGGCTGGGCATCTTGGTCAGCAAGTGGAGGAGCAGCAGTGCAGAGCGGGCTGGC 443
 Qy 40 LeuArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHis 59
 Db 444 CTGTGCTGGGGCAAGATCAAGAGGTGAATGGCTGAGCCTGGAGCAGCACCACTATG 503
 Qy 60 AlaGluAlaValLysAlaLysGlySerLysLysLeuValLeuSerValTySerAla 79
 Db 504 GGTAGCGCGTAAAGGTGTGACAGCAGCAGCGCCCTGCACATGATGGTTTGGCGCATG 563
 Qy 80 GlyArgIleProGlyGlyTyValThrAsnHisIleTyThrTrpValAspProGlnGly 99
 Db 564 GCGCGTGTCCGGGCATCAAGTTCCTCAAGGAGAGACCACTGGGTGGATGTGGTGAAT 623
 Qy 100 ArgSerIle-----SerProProSerGlyLeuProGlnProHisGly 113
 Db 624 CGCGCTGTGTAGTGGAGAAGTGGGTTCACACCTCCGAC----- 665
 Qy 114 GlyAlaLeuArgGlnGlnGlnGlyAspArgAspSerThrLeuHisLeuGlnGlyGly 133
 Db 666 -----ACCAGCTCAGAAGATGGTCCCGCGC---ATCGTCACTATACACACCTCC 716
 Qy 134 AspGluLysLysValAsnLeuValLeuGlyArgGlyAspGlyArgSerLeuGlyLeuThrIleArg 153
 Db 717 GACGAC-----TTCTGCTGGGCTTCAACATCCGT 746
 Qy 154 GlyGlyAlaGluTyGlyLeuGlyIleTyIleThrGlyValAspProGlySerGluAla 173
 Db 747 GGGGGCAAGAGTGTGGCTGGGCATCTATGTCTCAAAAGTGACCATGGTGGGTGGCC 806
 Qy 174 GluGlySerGlyLeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeu 193
 Db 807 GAGGAGATGGCATCAAGTGGGGGACCAAGTCTCGGACCAACCGTGTCAAGTTTGC 866
 Qy 194 AsnIleLeuHisAspGluAlaValArgLeuLeuLysSerSerArgHisIleLeuThr 213
 Db 867 GACATCAGCACACCGCAGCGGTGGAGGTGCTGAGGGGCAACCGCACATCATGCTGACC 926
 Qy 214 ValLysAspValGlyArgLeuProHisAlaArgThrThrValAspGluThrLysTrpIle 233
 Db 927 ATCAGAGAGACCGCGGTATCTGCTACAGGAGATGGTTCTGAGTACTGCTGGCTG 986
 Qy 234 -----AlaSerSerArgIleArgGluThrMetAlaAsnSerAlaGlyHisSer 251
 Db 987 GACCGACTGACCAACGGGTGCTGCAGCAGCTGTCCCGGCTCTGCAGAGCAGCTCCAGC 1046
 Qy 252 AlaArgSer 254
 Db 1047 GTCTCTTCG 1055

RESULT 5
 US-10-029-386-3728
 ; Sequence 3728, Application US/10029386
 ; Publication No. US20030194704A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharon G.
 ; APPLICANT: Rank, David R.
 ; APPLICANT: Hanzel, David K.
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
 ; FILE REFERENCE: AEOMICA-X-2
 ; CURRENT APPLICATION NUMBER: US/10/029,386
 ; CURRENT FILING DATE: 2001-12-20
 ; NUMBER OF SEQ ID NOS: 34288
 ; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
 ; SEQ ID NO 3728
 ; LENGTH: 509
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: MAP TO ALL38895.4
 ; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.63
 ; OTHER INFORMATION: NT HIT: G114735271, EVALUE 1.00e-123
 ; OTHER INFORMATION:

; OTHER INFORMATION: EST HUMAN HIT: BF969269.1, EVALUE 0.00e+00
 ; OTHER INFORMATION: SWISSPROT HIT: P54817, EVALUE 3.70e-02
 ; US-10-029-386-3728

Alignment Scores:
 Pred. No.: 1,99e-38 Length: 509
 Score: 408.50 Matches: 90
 Percent Similarity: 70.55% Conservative: 13
 Best Local Similarity: 61.64% Mismatches: 32
 Query Match: 30.81% Indels: 11
 DB: 15 Gaps: 2
 US-10-078-090-151 (1-260) x US-10-029-386-3728 (1-509)
 Qy 3 ArgAlaLysAlaHisGlnGlyLeuGlyPheSerIleArgGlyGlySerGluHisGlyVal 22
 Db 34 AAATCGAGGGCTAGAGAGGAAGGGA-----CTTCAGGGCTGG----- 72
 Qy 23 GlyIleTyValSerLeuValGluProGlySerLeuAlaGluLysGlnGlyLeuArgVal 42
 Db 73 -----CCTCTTCTTCCATGTGGGGCTCTCGAGAAAAACCAAGGTCTGGGCCAG 120
 Qy 43 GlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAlaGluAla 62
 Db 121 GGGAGGTGTCTTCCACACCCCGGCACCTCCCTTCCCTTAACCTTGTATCCCTCTCT 180
 Qy 63 ValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTySerAlaGlyArgIle 82
 Db 181 CCCTAGGCTCTGAGGGCTCCAAAGAGCTGGTGTCTGTGTACTCAGCAGGCGCGCATC 240
 Qy 83 ProGlyTyValThrAsnHisIleTyThrTrpValAspProGlnGlyArgSerIle 102
 Db 241 CCTGGGGGTCTACGTCACCAACCATCTACACCTGGGTGGACCCCGCAGGCGCCGAGCATC 300
 Qy 103 SerProProSerGlyLeuProGlnProHisGlyAlaLeuArgGlnGlnGlyAsp 122
 Db 301 TCCCCACCTCGGGCTGCCCGCCAGCCCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 360
 Qy 123 ArgArgSerThrLeuHisLeuGlnGlyGlyAspGluLysLysValAsnLeuValLeu 142
 Db 361 CGGAGGAGACCTCGACCTCTCTCAAGGAGGGGATGAGAAAAAGGTGATGGGTGGGG 420
 Qy 143 GlyAspGlyArgSerLeu 148
 Db 421 AAAGGAGCGCAGCTCTC 438

RESULT 6
 US-10-029-386-17428
 ; Sequence 17428, Application US/10029386
 ; Publication No. US20030194704A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharon G.
 ; APPLICANT: Rank, David R.
 ; APPLICANT: Hanzel, David K.
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
 ; FILE REFERENCE: AEOMICA-X-2
 ; CURRENT APPLICATION NUMBER: US/10/029,386
 ; CURRENT FILING DATE: 2001-12-20
 ; NUMBER OF SEQ ID NOS: 34288
 ; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
 ; SEQ ID NO 17428
 ; LENGTH: 195
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: MAP TO ALL38895.4
 ; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.63
 ; OTHER INFORMATION: EST HUMAN HIT: BE90168.1, EVALUE 1.00e-105
 ; OTHER INFORMATION: NT HIT: G114735271, EVALUE 1.00e-106
 ; OTHER INFORMATION: SWISSPROT HIT: P35428, EVALUE 3.70e-01
 ; US-10-029-386-17428

Alignment Scores:
Pred. No.: 1.15e-31 Length: 195
Score: 346.00 Matches: 65
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 26.09% Indels: 0
DB: 15 Gaps: 0

US-10-078-090-151 (1-260) x US-10-029-386-17428 (1-195)

QY 65 AlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAlaGlyArgIleProGly 84
Db 1 GCTCTGAAGGGCTCAAGAAGCTGCTGTCTGTGTACTCAGCAGGGCGCATCCCTGGG 60
QY 85 GlyTyrValThrAsnHisIleTyrThrTrpValAspProGlnGlyArgSerIleSerPro 104
Db 61 GGCTAGCTCACCACCAACATCTACACCTGGGTGGACCCGAGGGCGGAGCATCTCCCA 120
QY 105 ProSerGlyLeuProGlnProHisGlyAlaLeuArgGlnGlnGlyAspArg 124
Db 121 CCCTGGGGCTGCCCGAGCCCGCCGCTGCTGCTGAGGAGGAGGAGGAGGAGG 180
QY 125 SerThrLeuHisLeu 129
Db 181 AGCACCTGCACCTC 195

RESULT 7

US-10-177-191A-5
; Sequence 5, Application US/10177191A
; Publication No. US20030148381A1
; GENERAL INFORMATION:
; APPLICANT: Hybrigenics, Institut Pasteur
; APPLICANT: Legrain P.
; APPLICANT: David L.
; APPLICANT: Petit C.
; APPLICANT: Boeda B.
; APPLICANT: El-Amraoui A.
; TITLE OF INVENTION: More Protein-Protein Interactions In The Inner Bar
; FILE REFERENCE: B4940A
; CURRENT APPLICATION NUMBER: US/10/177,191A
; PRIOR FILING DATE: 2002-06-21
; PRIOR FILING DATE: 2001/06/21
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patent version 3.1
; SEQ ID NO 5
; LENGTH: 837
; TYPE: DNA
; ORGANISM: Human
US-10-177-191A-5

Alignment Scores:
Pred. No.: 8.07e-27 Length: 837
Score: 312.00 Matches: 83
Percent Similarity: 51.53% Conservative: 35
Best Local Similarity: 36.24% Mismatches: 83
Query Match: 23.53% Indels: 28
DB: 15 Gaps: 6

US-10-078-090-151 (1-260) x US-10-177-191A-5 (1-837)

QY 8 GluGlyLeuGlyPheSerIleArgGlyGlySerGluHisGlyValGlyIleTyrValSer 27
Db 19 GAAGGTCTCGGCTCAGCGTGGTGGAGGCGCTGGAATTTGGCTGTGACTCTTTATCTCC 78
QY 28 LeuValGluProGlySerLeuAlaGluLysGluGlyLeuArgValGlyAspGlnIleLeu 47
Db 79 CACCTCAACAAAGTGGCCAGCAGAGCGGCTTGGCTTTCAGGTAGGGGATGAATGTC 138
QY 48 ArgValAsnAspLysSerLeuAlaArgValThrHisAlaGluAlaValLysAlaLeuLys 67
Db 139 CGGATCAAGGCTATTCCATCTCTCTCTGTAACCATGAGGAAGTCATCAACCTGATCCGC 198

QY 68 GlySerLysLysLeuValLeuSerValTyrSerAlaGlyArgIleProGlyGlyTyrVal 87
Db 199 ACCAAGAAGACCGTGTCCATCAAGTGAGACACATCGGACTGATCCCTGTGAAGACTCT 258
QY 88 ThrAsnHisIleTyrThrTrp-----ValAspProGlnGlyArgSerIleSerProPro 105
Db 259 CTTGAGGAGTCCCTCAATGGCAGTATGTGGAT-----CAGTTCGTGCGAATCT 309
QY 106 SerGlyLeuProGlnProHisGlyAlaLeuArgGlnGlnGluGlyAspArgArgSer 125
Db 310 GGGGGTGTG-----CGAGGTGGCTTG-----GGCTCACCTGGCAATCGGACAACC 354
QY 126 ThrLeuHisLeuLeuGlnGlyAspGlnLysLysValAsnLeuValLeuGlyAspGly 145
Db 355 -----AAGGAGAAAGAGGTGTTTATCATGCTAGTGGGCTCT 390
QY 146 ArgSerLeuGlyLeuThrIleArgGlyGlyAlaGluTyrGlyLeuGlyIleTyrIleThr 165
Db 391 CGGGGCTGGGTGAGCATCTCCAGTGGCCCATCCAGAGCGCTGCATCTTCGTGACG 450
QY 166 GlyValAspProGlySerGluAlaGluGlySerGlyLeuLysValGlyAspGlnIleLeu 185
Db 451 CAGGTGAAGCTGGCTCCCTGCTGTCAGAGGTGGGGTTAGAGACAGGACAGACAGATTGTG 510
QY 186 GluValAsnGlyArgSerPheLeuAsnIleLeuHisAspGluAlaValArgLeuLys 205
Db 511 GAAGTCAATGGCATAGACTTCACCACTGGAGCCCAAGGAGGCTGTGAATGCTCTGAAG 570
QY 206 SerSerArgHisLeuIleLeuThrVal-----Lys 215
Db 571 AGCAGCGCAGCGCTGACCATCTCCATGCTGTGAGAGCGCGGAGCTGTTTCATGACG 630
QY 216 AspValGlyArgLeuProHisAlaArg 224
Db 631 GACCGGAACGGCTGGAGGAGGACCGG 657

RESULT 8

US-10-305-720-1138
; Sequence 1138, Application US/10305720
; Publication No. US20040010136A1
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice K.; Seilhamer, Jeffrey J.
; TITLE OF INVENTION: Composition for the Detection of Signaling Pathway Gene Expressic
; FILE REFERENCE: PA-0002-1 CON
; CURRENT APPLICATION NUMBER: US/10/305,720
; CURRENT FILING DATE: 2002-11-26
; PRIOR FILING DATE: 1998-01-30
; NUMBER OF SEQ ID NOS: 1490
; SOFTWARE: PERL Program
; SEQ ID NO 1138
; LENGTH: 3071
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: GenBank ID No. US20040010136A1 g1463025
US-10-305-720-1138

Alignment Scores:
Pred. No.: 2.55e-16 Length: 3071
Score: 229.00 Matches: 104
Percent Similarity: 39.89% Conservative: 38
Best Local Similarity: 29.21% Mismatches: 90
Query Match: 17.27% Indels: 124
DB: 16 Gaps: 19

US-10-078-090-151 (1-260) x US-10-305-720-1138 (1-3071)

QY 5 LysAlaHisGluGlyLeuGlyPheSerIleArgGly-----GlySerGluHisGlyVal 22
Db 868 AAAGCCCTAAAGGTTTAGGCTTCAGTATTCAGGAGGTGGGGAACCAACACATCTCT 927

QY 23 Gly-----IleTyrValSerLeuValGluProGlySerLeuAlaGluGly 39
Db 928 GGGGCAACAGCATTTATGTAACAAATATTAGATGGAGGCTGCACAAAGATGGA 987
QY 40 ----LeuArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThr 58
Db 988 AGGTTGCAAGTAGGAGATAGACTACTAATGTTAAACAACTACAGTTTAGAAGAGTAA 1047
QY 59 HisAlaGluAlaValIleAlaLeuLysGlySerLysLeuVal---LeuSerValTyr 77
Db 1048 CACGAGAGGAGTAGCANTATTAAAGACACATCAGAGGTAGTTATTAAAGATTGGC 1107
QY 78 SerAlaGlyArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTyrValAspPro 97
Db 1108 AACCCCACTACCATTT---TATATGACT---GATCCT 1137
QY 98 GlnGly-----ArgSerIleSerProPro-----Ser 106
Db 1138 TAGGTCCCACTGATATTACTACCTCTTATTCTCCACCAATGAAACCACTACTCTCT 1197
QY 107 Gly----- 107
Db 1198 GGCACAATGGCACTTTAGAAATATAAACTCCCTGCCACCATCTCTCCAGGGAGGTAC 1257
QY 108 ----LeuProGlnPro--- 111
Db 1258 TCACCAATTCCAAAGCACATGCTTTGAGGAGCAGTACACAGGCTCCGAACTGTT 1317
QY 111 ----- 111
Db 1318 TACGACCTGTGAACAACTATGTGATAAGCCTGTTCTCCAGGCACTATTCCCTGTT 1377
QY 112 -----HisGlyGlyAlaLeu 116
Db 1378 GAGTGTGACAAAGCTTCCTCCTCAGCTCCCTATTCCCATACCATGAGCTGCTA 1437
QY 117 ArgGlnGlnGlu-----GlyAspArgArgSerThrLeu 127
Db 1438 CCTGACTCTGAGTACCACTGATCCCAATAGCACGCACTCGTCAGCCTTCAATG 1497
QY 128 HisLeuLeuGlnGly-----GlyAspGluLysIleValAsnLeuValLeuGly 143
Db 1498 ACTCTCCAAAGCGGCGCTCCCTGGAGGAGAGCCTCGCAAGGTATCTGTCACAAAGC 1557
QY 144 AspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGluTyrGlyLeuGlyIleTyr 163
Db 1558 ---TCCACTGGCTGGGCTTCAACATTTGCGTGGG---GAGATGGAGAAGGTATTTT 1611
QY 164 IleThrGlyValAspProGlySerGluAlaGluGlySerGly---LeuLysValGlyAsp 182
Db 1612 GTGTCTTCAATTCGTGGTGGTGGACCACTAAGTGGGAGCTCCAGAGAGAGAC 1671
QY 183 GlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHisAspGluAlaValArg 202
Db 1672 CAGATCTTATCGTGATGCAATGATGCTCCGTGGTGCATCCACAGCAGGAGCTGCT 1731
QY 203 LeuLeuLysSerArgHisLeuIleLeuThrValLysAspValGlyArgLeuProHis 222
Db 1732 GCACCTAAAGGGGCTGGACACACAGTACGATATATACCAATATCAACCTGAAGATTAC 1791
QY 223 AlaArgThrThrValAspGluThrLysTrpIleAlaSerSerArgIleArgGluThrMet 242
Db 1792 GCTCGATTT-----GAGGCCAA-----ATCATGACCTACAGAGCAGATG 1833
QY 243 AlaAsn-----SerAlaGlySerGlyHisSerAlaArgSerAsn 255
Db 1834 ATGAACCAACAGATGAGCTCGGGTCCGGA---TCCCTGGCAACCAAT 1878

RESULT 9
US-10-451-207-14
; Sequence 14, Application US/10451207
; Publication No. US20040038267A1
; GENERAL INFORMATION:

APPLICANT: INCYTE GENOMICS, INC.
APPLICANT: BURFORD, Neil
APPLICANT: DING, Li
APPLICANT: YUE, Henry
APPLICANT: THORNTON, Michael B.
APPLICANT: CHAWLA, Navinder K.
APPLICANT: GANDHI, Ameena R.
APPLICANT: ARVIZU, Chandra S.
APPLICANT: BAUGHN, Mariah R.
APPLICANT: SWARNAKAR, Anita
APPLICANT: DUGGAN, Brendan M.
APPLICANT: LU, Dying Aina M.
APPLICANT: THANGAVELU, Kavitha
APPLICANT: WARREN, Bridget A.
APPLICANT: TANG, Y. Tom
APPLICANT: KHAN, Farrah A.
APPLICANT: YAO, Monique G.
APPLICANT: EMERLING, Brooke M.
TITLE OF INVENTION: INTRACELLULAR SIGNALING MOLECULES
FILE REFERENCE: PF-0868 USN
CURRENT FILING DATE: 2003-06-18
CURRENT APPLICATION NUMBER: US/10/451,207
PRIOR APPLICATION NUMBER: PCT/US01/50315
PRIOR FILING DATE: 2001-12-19
PRIOR APPLICATION NUMBER: US 60/257,804
PRIOR FILING DATE: 2000-12-21
PRIOR APPLICATION NUMBER: US 60/260,102
PRIOR FILING DATE: 2001-01-05
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PERL Program
SEQ ID NO 14
LENGTH: 5139
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20040038267A1 6942051CB1
US-10-451-207-14
Alignment Scores:
Pred. No.: 1,01e-15 Length: 5139
Score: 226.50 Matches: 82
Percent Similarity: 42.96% Conservative: 37
Best Local Similarity: 29.60% Mismatches: 81
Query Match: 17.08% Indels: 77
DB: 13 Gaps: 10
US-10-078-090-151 (1-260) x US-10-451-207-14 (1-5139)
QY 1 LeuArgArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGlyGlySerGluHis 20
Db 3013 ATCCCTCTGCCAAGAGCTGGGGGCCCTCTGGGGCTTAGTATTGTCGGAGGCTCCGACCAT 3072
QY 21 -----GlyVal-----GlyIleTyrValSerLeuValGluProGly 32
Db 3073 TCCAGCCACCCGCTTGTGTCTCCAGGAGCCTGCTGTTCATCTCCAAGTCTCCCGCG 3132
QY 33 SerLeuAlaGluLysGluGlyLeuArgValGlyAspGlnIleLeuArgValAsnAspLys 52
Db 3133 GGCCTTCGCGCTCCGAGCGGCTCGGGTTGGGACCCCATCTCTGGCAGTGAACGGGCAA 3192
QY 53 SerLeuAlaArgValThrHisAlaGluAlaValLysAla---LeuLysGlySerLysLys 71
Db 3193 GACGTGCGGATGCCACACGACCAAGACAGTCAGTGCCTGCTCCGCGCCTGCTGGAG 3252
QY 72 LeuValLeuSerValTyrSerAlaGlyArgIleProGlyGlyTyrValThrAsnHisIle 91
Db 3253 CTGTCGCTGCTGGTG----- 3267
QY 92 TyrThrTrpValAspProGlnGlyArgSerIleSerProProSerGlyLeuProGlnPro 111
Db 3268 -----CGGAGGGACCCCGGACCCCGCGGCTA----- 3294

```

QY 112 HisGlyGlyAlaLeuArgGlnGlnGluGlyAspArgSerThrLeuHisLeuLeuGln 131
Db 3294 -----
QY 132 GlyGlyAspGluLysLysValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThr 151
Db 3295 -----
QY 152 IleArgGlyGlyAlaGlu-----TyrGlyLeuGlyLeu 162
Db 3343 ATCCGCGGGGTGCAGGGGCCACGGTGGCAACCCCGCGACCCACAGAGAGGGGATC 3402
QY 163 TyrIleThrGlyValAspPro---GlySerGluAlaGluGlySerGlyLeuLysValGly 181
Db 3403 TTCATCTCAAGGTAGAGCCCGGCGGACCCCGGCGGCGGCGGCGGCGGCGGCGG 3462
QY 182 AspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHisAspGluAlaVal 201
Db 3463 TTGGGCTGTTGGAGGTGAACAGCAGAGAGCTGTGGGCTGAGCGACGCGGCGGCGG 3522
QY 202 ArgLeuLeuLysSerSerArgHisLeuLeuThrValLysAspValGlyArgLeuPro 221
Db 3523 CAGCTGCTCCGAGGTGGGCGGACACCTCACCGTGTGGTCTGTGACGGC----- 3573
QY 222 HisAlaArgThrValAspGluThrLysTrpIleAlaSerSerArgIleArgGluThr 241
Db 3574 ---TTCGAGGCCAGCAGCAGCCCGTGGAGGTGTCCCGAGGTGTCTATTGCCAACCCC 3630
QY 242 MetAlaSerAlaGlySerGlyHisSerAlaArgSerAsnLeuGlnThr 258
Db 3631 TTTTGGC-----GCAGGATCGGCCAC-----CGGAACAGCTGGAGAGC 3669

RESULT 10
US-09-919-497-9
; Sequence 9, Application US/09919497
; Patent No. US2002010662A1
; GENERAL INFORMATION:
; APPLICANT: Muttter, George L.
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
; CURRENT APPLICATION NUMBER: US/09/919,497
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/221,735
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 3995
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-919-497-9

Alignment Scores:
Pred. No.: 2,42e-15 Length: 3995
Score: 222.00 Matches: 91
Percent Similarity: 46.45% Conservativity: 40
Best Local Similarity: 32.27% Mismatches: 110
Query Match: 16.74% Indels: 41
DB: Gaps: 13

US-10-078-090-151 (1-260) x US-09-919-497-9 (1-3995)
QY 1 LeuArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGly-----GlySer 18
Db 1469 ATCAAGCTCATCAGGGGCCCTAAGGTCTTGGCTTACGATCGCAGGGGCGGTAGGAGAC 1528
QY 19 GluHisGlyValGly-----IleTyrValSerLeuValGluProGlySerLeuAla 35
Db 1529 CAGCACATCCAGGAGATAATAGCATCTATGTAAACAAGATCATCGAAGGGGTGTCTGCC 1588
QY 36 GluLysGluGly---IleuArgValGlyAspGlnIleuArgValAsnAspLysSerLeu 54
Db 1589 CACAAGCATGGAGGTTCAGATTGGAGACAAGATCTCTCGCGGTCAACAGTGTGGGGCTA 1648

```

```

QY 55 AlaArgValThrHisAlaGluAlaValLysAlaLeuLysGlySerLysLeuValLeu 74
Db 1649 GAGGAGCTATGATCATGAAGTGTCTGTGCAGCCCTGAAGAACACGATGATGTCTCTAC 1708
QY 75 -----SerValTyrSerAlaGlyArgIleProGlyGlyTyr 86
Db 1709 CTAAGGTGGCCAGCCAGCAATGCCTACCTAGTACAGCTATGCTCCCCAGACATC 1768
QY 87 ValThrAsnHisIleTyr-ThrTrpValAspProGlnGlyArgSerIleSerProSe 106
Db 1769 ACAACCTCTATTCCAGCACCTGGACA-----ATGAGATCAGTCACAGCAGCTAC 1819
QY 106 rGlyLeuProGln-ProHis-----GlyGlyAlaLeuArgGlnG 119
Db 1820 CTGGGACCCGATACCCACAGCCATGACCCCACTTCCCTCGGCGGTACTCTCCAGTG 1879
QY 119 InGluGlyAspArgArgSerThrLeuHisLeuGlnGlyAspGluLysVala 139
Db 1880 GCCAAGGACCTCTCGGGAGGAGAACAT-----TCCCGGAGAACCCGAGGCGA 1927
QY 139 sn-LeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGluTyr 158
Db 1928 ATGTGATCCACCGGGGCTCCACGGGCTTGGCTTCAACATCTGTGGTGGC---GAGGAC 1984
QY 159 GlyLeuGlyIleTyrIleThrGlyValAspProGlySerGluAlaGluGlySerGly 177
Db 1985 GGTGAAGGATCTTCACTCTCTTTATCTCGCGGGGCGCCCTGCAGACCTCAGTGGGAG 2044
QY 178 LeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHis 197
Db 2045 CTCGGAAGGGGACCCAGATCTGTCTGGTCAACGGTGTGGACCTCCGAAATGCCAGCCAT 2104
QY 198 AspGluAlaValArgLeuLeuLysSerArgHisLeuLeuLeuThrValLysAspVal 217
Db 2105 GAGCAGGCTGCCATTCCTGCTGAAGAATGCGGGTCAAGCGTACAGATCATCGCTCAGTAT 2164
QY 218 GlyArgLeuProHisAlaArgThrValAspGluThrLysTrpIleAlaSerSerArg 237
Db 2165 AACCCAGAGTACAGCCGATTC-----GAGCCCAAG-----ATCCACGAC 2206
QY 238 IleArgGluThrMetAlaAsnSerAla---GlySerGly---HisSerAlaArgSerAsn 255
Db 2207 CTTCCGGAACAGCTCATGACAGCAGCTGGGCTCAGGAGCTCGTCTTGGCGGAGCAAC 2266

RESULT 11
US-10-170-385-300
; Sequence 300, Application US/10170385
; Publication No. US2003020372A1
; GENERAL INFORMATION:
; APPLICANT: Ward, Neil Raymond
; APPLICANT: Mundy, Christopher Robert
; APPLICANT: Kan, On
; APPLICANT: Harris, Robert Alan
; APPLICANT: White, Jonathan
; APPLICANT: Binley, Katie Mary
; APPLICANT: Rayner, William Nigel
; APPLICANT: Naylor, Stuart
; APPLICANT: Kingsman, Susan Mary
; APPLICANT: Krige, David
; TITLE OF INVENTION: ANALYSIS METHOD
; FILE REFERENCE: 53268200100
; CURRENT APPLICATION NUMBER: US/10/170,385
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: PCT/GB02/01662
; PRIOR FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: PCT/GB01/05458
; PRIOR FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 549
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 300
; LENGTH: 3995
; TYPE: DNA

```


QY 193 LeuAsnIleLeuHisAspGluAlaValArgLeuLeuLysSerArgHisLeuLeu 212
Db 2705 CAGAAC-----CGTGAAGAGCGTGTGGCTCTTCTAACCACTGAGAAATAAAACTTT 2652
QY 213 ThrValLysAspValGlyArgLeuProHisAlaArgThrValAspGluThrLysTrp 232
Db 2651 TCATTG-----CTGATTCGAAGGCTGAACCTCCAGCTGGATGAG---GGCTGG 2607
QY 233 IleAlaSerArg 237
Db 2606 ATGGATGATGACAGG 2592
RESULT 13
US-10-296-115-496/c
; Sequence 496, Application US/10296115
; Publication No. US20040053248A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq Inc
; TITLE OF INVENTION: No. US20040053248A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 784PCT
; CURRENT APPLICATION NUMBER: US/10/296,115
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: US09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US09/552,317
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 1478
; SEQ ID NO 496
; LENGTH: 4235
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-296-115-496
Alignment Scores:
Pred. No.: 4,42e-13 Length: 4235
Score: 203.00 Matches: 75
Percent Similarity: 46.42% Conservative: 48
Best Local Similarity: 28.30% Mismatches: 98
Query Match: 15.31% Indels: 44
DB: 13 Gaps: 12
US-10-078-090-151 (1-260) x US-10-296-115-496 (1-4235)
QY 10 LeuGlyPheSerIleArgGlyGly-----SerGluHisGly 21
Db 3365 CTGGGATTCAATATTATTGTTGGCGCGCGGTGGTATACACGATGATGATCCAGT 3306
QY 22 ValGlyIleTrpValSer---LeuValGluProGlySerLeuAlaGluLysGluGlyLeu 40
Db 3305 GAAGGAATCTTGTATCTCAAGATAGTTGACAGTGGCGCTGCACGCAAGAGAGGCGCTG 3246
QY 41 ArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAla 60
Db 3245 CAATTCATGACAGGATTATTGAGTCAACGCGCAGAGACTTTATCCAGACCACTCATGAC 3186
QY 61 GluAlaValLysAlaLeuLysGlySerLysLys---LeuValLeuSerValTrp----- 77
Db 3185 CAGGCTGTGGAAGCTTTCAAGACAGCCAGAGGCCATAGTGGTGGTGTGAGAGA 3126
QY 78 -----SerAlaGlyArgTleProGlyGlyTrp 86
Db 3125 ACACCAAGACCAAAATGTTCCAGCTCCATCAGACTTCACCTGGTGGACAGCGAACC 3065
QY 87 ValThrAsnHisIleTrpThrTrpValAspProGlnGlyArgSerIleSerPro----- 104
Db 3065 CAAACCGACATCACCTTTGAACATATCATGGCCCTCCTCACTAAGATGCTCTCCAGCCCA 3006
QY 105 -----ProSerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlu 120
Db 3005 CCGTGCTGATCCCTATCTCTTCAGAGAGCATCCCTCAGCCCATGAACTACCAT 2946
QY 121 ---GlyAspArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysValAsn 139
Db 121 ---GlyAspArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysValAsn 139

Db 2945 CCAATGACTACATTGGAGACATCCATCAGGAGATGGACAGGAGAGCTGGAGCTGGAG 2886
QY 140 LeuVal-----LeuGlyAspGlyArgSerLeuGlyLeuThrIle-----Arg 153
Db 2885 GAATGGAGCTCTACAGAATGAACAGCCAGGACAGCTGGGCTCCTCTGTGTCTACCG 2826
QY 154 GlyGlyAlaGluTrpGlyLeuGlyIleTrpIleThrGlyValAspProGlySerGluAla 173
Db 2825 ACGGACGATGAAGACGACATTTGGATTTATATCAGTGAGATTGACCTAACAGCATTCGA 2766
QY 174 GluGlySerGly---LeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPhe 192
Db 2765 GCCAAGATGGCGCATCCGAGAGAGACCGCATTCATCCAGATTATGGTAGAGGTG 2706
QY 193 LeuAsnIleLeuHisAspGluAlaValArgLeuLeuLysSerArgHisLeuLeu 212
Db 2705 CAGAAC-----CGTGAAGAGCGTGTGGCTCTTCTAACCACTGAGAAATAAAACTTT 2652
QY 213 ThrValLysAspValGlyArgLeuProHisAlaArgThrValAspGluThrLysTrp 232
Db 2651 TCATTG-----CTGATTCGAAGGCTGAACCTCCAGCTGGATGAG---GGCTGG 2607
QY 233 IleAlaSerArg 237
Db 2606 ATGGATGATGACAGG 2592
RESULT 14
US-09-964-956-24
; Sequence 24, Application US/09964956
; Publication No. US20040043926A1
; GENERAL INFORMATION:
; APPLICANT: Gerlach, Valerie L
; APPLICANT: MacDougall, John R
; APPLICANT: Smithson, Glennda
; APPLICANT: Millet, Isabelle
; APPLICANT: Stone, David
; APPLICANT: Gunther, Erik
; APPLICANT: Ellerman, Karen
; APPLICANT: Grosse, William M
; APPLICANT: Alsobrook II, John P
; APPLICANT: Lepley, Denise M
; APPLICANT: Burgess, Catherine E
; APPLICANT: Padigaru, Muraidhara
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Leach, Martin D
; APPLICANT: Shimkets, Richard A
; TITLE OF INVENTION: No. US20040043926A1el Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-124
; CURRENT APPLICATION NUMBER: US/09/964,956
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/235,631
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/235,633
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/235,808
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/236,064
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/236,065
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/236,066
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/236,135
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: 60/237,434
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/238,321
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/238,399
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/238,396
; PRIOR FILING DATE: 2000-10-06

; PRIOR APPLICATION NUMBER: 60/276,667
 ; PRIOR FILING DATE: 2001-03-16
 ; PRIOR APPLICATION NUMBER: 60/294,823
 ; PRIOR FILING DATE: 2001-05-31
 ; PRIOR APPLICATION NUMBER: 60/304,868
 ; PRIOR FILING DATE: 2001-07-12
 ; NUMBER OF SEQ ID NOS: 127
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 24
 ; LENGTH: 8640
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-09-964-956-24

Alignment Scores:

Pred. No.: 4,518-12 Length: 8640
 Score: 198.00 Matches: 87
 Percent Similarity: 36.72% Conservative: 43
 Best Local Similarity: 24.58% Mismatches: 124
 Query Match: 14.93% Indels: 100
 DB: 13 Gaps: 13

US-10-078-090-151 (1-260) x US-09-964-956-24 (1-8640)

QY 5 LysAlaHisGluGlyLeuGlyPheSerIleArgGlyGly-----SerGlu 19
 DB 1121 AAAGAATCGATGGCTGGGAATTCAGTTAGTGGAGGCCGAGATCAAGCGCTCACT 1180
 QY 20 HisGlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGlu-----Lys 37
 DB 1181 CAC-----GCTATCGTTGCTCACTCAAGTGAAGAGAGGAGGCGCTCACAGGCTCAGG 1234
 QY 38 GluGly-----LeuArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArg 56
 DB 1235 GATGCGAGCGTGTCTTAGAGATGAGTGTCTGTAATCAATGGTCAATTTACTGTGCGG 1294
 QY 57 ValThrHisAlaGluAlaValLysAlaLeuLysGlySerLysLysLeuVal----- 73
 DB 1295 CTCTCCACAGGAGAGAGTGGCCATCTTCGCTCCGCCACGGATGTGCGAGTTGTG 1354
 QY 74 -----LeuSerValTyrSerAlaGlyArgIleProGlyGlyTyrValThrAsn 89
 DB 1355 GTGCCACGAGGATAGTGTGTCTTCTGCAATTCAGATGCTGGACAGATGAACCCCAA 1414
 QY 90 HisIleTyrThrIleValAspProGlnGlyArgSerIleSerProSerGly----- 107
 DB 1415 GATGTGTGCGGTGCTAGGAATCCAGGGGAACCTTGAAGAGTCCCAACAGGCGACCAAT 1474
 QY 108 -----LeuProGlnProHis 112
 DB 1475 AAATCAAGCTCAAGATCGCTTTTACAGTGTAGTGGGGCTCTACCTGTGATGAGCCTGTC 1534
 QY 113 GlyGlyAlaLeuArg----- 117
 DB 1535 GGGGGGTACACCGCTTGTAGTCAATGTAAGATATTAACAGAGTGTGTCGGATGGG 1594
 QY 118 -----GlnGlnGlyAspArgArgSerThrLeuHis 128
 DB 1595 GACCCCGGATCCGGATGTTGGAGGTCTCCGAGATGCGCGAAACACTCCCTCCGCGAG 1654
 QY 129 LeuLeuGlnGlyAspGluLysLys----- 137
 DB 1655 CTGTGGACTCTTCCAGTGCCTCAGAGATACACATTTGTGAAGAAGTCTACCGGCTCC 1714
 QY 137 ----- 137
 DB 1715 TTAACGACACTAGGTGATCTCTTGGAGGCTATTCGGCCATCCGTCATCTCGATC 1774
 QY 138 ValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGlu 157
 DB 1775 ATTGGTGTACAAAGAAAGGCAAGCGCTTGTAGTATTGTGAGGTCGAGAC 1834
 QY 158 Tyr-----GlyLeuGlyIleTyrIleThrGlyValAspPro---GlySerGluAla 173

DB 1835 TGCATTCGTGGACAGATGGGATTTTGTCAAGACCATCTTCCCAATGGATCAGTCCA 1894
 QY 174 GluGlySerGlyLeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeu 193
 DB 1895 GAGGACGGAAGACTTAAGAAGGTGATGAATCTTAGATGTAATGGAATACCAATAAG 1954
 QY 194 AsnIleLeuHisAspGluAlaValArgLeuLeuLysSerSerArgHis-----LeuIle 211
 DB 1955 GGCTTCACATTTCAAGAGCCATTCATACCTTTAAGCAAAATCCGGAGTGGATTATTGTT 2014
 QY 212 LeuThrValLysAspValGlyArgLeuProHisAlaArg-----ThrThrValAspGlu 229
 DB 2015 TTAACGGTAGCACAAGAGTGTGAGCCCGACGCTCACACCTGTCTGCACCCACACAC 2074
 QY 230 ThrLysTrpIleAlaSerSerArgIleArgGluThrMetAlaAsnSerLagly----- 247
 DB 2075 ATGAGCAGATCCGCTCCCGAACTTCAATACAGTGGGGAGCTTCGCGGAGGTTC 2134
 QY 248 ---SerGlyHisSerAlaArgSerAsnLeuGlnThrProGly 260
 DB 2135 GATGAAGGCAGTCTTCTCATCCCTGGTGGTGGAGAGACCCCTGGG 2176

RESULT 15
 US-09-964-956-26
 ; Sequence 26 Application US/09964956
 ; Publication NO. US20040043926A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Gerlach, Valerie R
 ; APPLICANT: MacDougall, John R
 ; APPLICANT: Smithson, Glenda
 ; APPLICANT: Millet, Isabelle
 ; APPLICANT: Stone, David
 ; APPLICANT: Gunther, Erik
 ; APPLICANT: Ellerman, Karen
 ; APPLICANT: Grosse, William M
 ; APPLICANT: Alsbrook II, John P
 ; APPLICANT: Lepley, Denise M
 ; APPLICANT: Burgess, Catherine E
 ; APPLICANT: Padigaru, Muralidhara
 ; APPLICANT: Kekuda, Ramesh
 ; APPLICANT: Spytek, Kimberly A
 ; APPLICANT: Leach, Martin D
 ; APPLICANT: Shinkets, Richard A
 ; TITLE OF INVENTION: No. US20040043926A1el Proteins and Nucleic Acids Encoding Same
 ; FILE REFERENCE: 21402-124
 ; CURRENT APPLICATION NUMBER: US/09/964,956
 ; CURRENT FILING DATE: 2001-09-26
 ; PRIOR APPLICATION NUMBER: 60/235,631
 ; PRIOR FILING DATE: 2000-09-27
 ; PRIOR APPLICATION NUMBER: 60/235,633
 ; PRIOR FILING DATE: 2000-09-27
 ; PRIOR APPLICATION NUMBER: 60/235,808
 ; PRIOR FILING DATE: 2000-09-27
 ; PRIOR APPLICATION NUMBER: 60/236,064
 ; PRIOR FILING DATE: 2000-09-27
 ; PRIOR APPLICATION NUMBER: 60/236,065
 ; PRIOR FILING DATE: 2000-09-27
 ; PRIOR APPLICATION NUMBER: 60/236,066
 ; PRIOR FILING DATE: 2000-09-27
 ; PRIOR APPLICATION NUMBER: 60/236,135
 ; PRIOR FILING DATE: 2000-09-28
 ; PRIOR APPLICATION NUMBER: 60/237,434
 ; PRIOR FILING DATE: 2000-10-03
 ; PRIOR APPLICATION NUMBER: 60/238,321
 ; PRIOR FILING DATE: 2000-10-05
 ; PRIOR APPLICATION NUMBER: 60/238,399
 ; PRIOR FILING DATE: 2000-10-06
 ; PRIOR APPLICATION NUMBER: 60/238,396
 ; PRIOR FILING DATE: 2000-10-06
 ; PRIOR APPLICATION NUMBER: 60/276,667
 ; PRIOR FILING DATE: 2001-03-16
 ; PRIOR APPLICATION NUMBER: 60/294,823

; PRIOR FILING DATE: 2001-05-31
 ; PRIOR APPLICATION NUMBER: 60/304,868
 ; PRIOR FILING DATE: 2001-07-12
 ; NUMBER OF SEQ ID NOS: 127
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 26
 ; LENGTH: 8640
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-09-964-956-26

Alignment Scores:
 Pred. No.: 4,51e-12 Length: 8640
 Score: 198.00 Matches: 87
 Percent Similarity: 36.72% Conservative: 43
 Best Local Similarity: 24.58% Mismatches: 124
 Query Match: 14.93% Indels: 100
 DB: 13 Gaps: 13

US-10-078-090-151 (1-260) x US-09-964-956-26 (1-8640)

```

QY      5 LysAlaHisGluGlyLeuGlyPheSerIleA-gGlyGly-----SerGlu 19
Db      1121 AAAGATCGATGGCTGGGAATTCAGTTAGTGGAGCGGAGGATCAAGGCTCACCT 1180

QY      20 HisGlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGlu-----Lys 37
Db      1181 CAC-----GCTATCGTTGCTCACTCAAGTGAAGGAGGAGGTCGCCGCTCACAGGCTCAGG 1234

QY      38 GluGly---LeuArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArg 56
Db      1235 GATGGCAGGCTGCTTCCTAGAGATGAGCTGCTGGTGAATCAATGGTCATTACTGCTCGGG 1294

QY      57 ValThrHisAlaGluAlaValLysAlaLeuLysGlySerLysLeuVal-----73
Db      1295 CTCTCCACAGGAGGAGTGGCCATTCTTCGTCGCCACCGGAATGTCGAGCTTGTG 1354

QY      74 -----LeuSerValTyrSerAlaGlyArgIleProGlyGlyTyrValThrAsn 89
Db      1355 GTGCCAGCAGGTAGTGTGCTTCTGCAATTCAGATTCCTGGACAGATGAACCCCA 1414

QY      90 HisIleTyrThrTrpValAppProGlnGlyArgSerIleSerProSerGly-----107
Db      1415 GATGTGTGGCGTCTGAGGAATCCAAAGGGGAAGTTCGAAAGTCCCAACAGAGGAGCAAT 1474

QY      108 -----LeuProGlnProHis 112
Db      1475 AAATCAAGCTCAAGAGTCGCTTTCAGGTAGTGGGGCTCTACCTGATGACGCTGTC 1534

QY      113 GlyGlyAlaLeuArg-----117
Db      1535 GGGGTGTACACCGCTTGAGTCAGTTGAAGATATAACGAGCTGATGTGCGGAATGGG 1594

QY      118 -----GlnGlnGluGlyAspArgArgSerThrLeuHis 128
Db      1595 GACCCCGGATCCGGATGTTGGAGGTCTCCCGAGATGGCGGAACACTCCCTCCCGCAG 1654

QY      129 LeuLeuGlnGlyGlyAspGluLysLys-----137
Db      1655 CTGCTGACTCTTCCAGTGGCTCACAGGAATACCACATTGTGAAGAGTCTACCCGCTCC 1714

QY      137 -----137
Db      1715 TTAAGCAGCACTCAGGTGAATCTCTCGGAGGCTCATTCGGCCATCCGTCATCTCGATC 1774

QY      138 ValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGlu 157
Db      1775 ATTGGGTGTACAAGAAAGGAGGAGGCTTGGCTTTAGTATTGCTGGAGGTCGAGAC 1834

QY      158 Tyr-----GlyLeuGlyIleTyrIleThrGlyValAspPro---GlySerGluAla 173
Db      1835 TGCATTCTGTGGACAGATGGGGATTTTGTCAAGACCATTCTCCCAATGGATCAGCTGCA 1894
  
```

```

QY      174 GluGlySerGlyLeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeu 193
Db      1895 GAGGACGGAAGACTTAAAGAAAGGGGATGAATCTAGATGTAATGGAATACCAATAAAG 1954

QY      194 AsnIleLeuHisAspGluAlaValArgLeuLeuLysSerSerArgHis-----LeuIle 211
Db      1955 GGCTTGACATTTCAGAAGGCCATTCTACCTTTAAGCAATCCGGAGTGGATTATTGTT 2014

QY      212 LeuThrValLysAspValGlyArgLeuProHisAlaArg-----ThrThrValAspGlu 229
Db      2015 TTAACGGTACGCACAAAAGTTGGTGAGCCCGAGGCTCACACCCCTGCTCGACACCCACACAC 2074

QY      230 ThrIleTrpIleAlaSerSerArgIleArgGluThrMetAlaAsnSerAlaGly-----247
Db      2075 ATGAGCAGATCCCGCTCCCGCAACTTCAATACCAGTGGGGGAGCCTCGGCGGAGGTTCC 2134

QY      248 ---SerGlyHisSerAlaArgSerAsnLeuGlnThrProGly 260
Db      2135 GATGAAGGCAGTCTTTCATCCCTGGGTGCGAAGACCCCTGGG 2176
  
```

Search completed: April 22, 2004, 03:48:24
 Job time : 407 secs

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 22, 2004, 00:07:22 ; Search time 2513 Seconds
(without alignments)
3089.601 Million cell updates/sec

Title: US-10-078-090-151
Perfect score: 1326
Sequence: 1 LRRKAHEGLGFSIRGSGH.....TMSAGSGHSARSLQTPG 260

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n_model -DEV=xlp
-O/cn2_1/USPTO_spool_p/US10078090/runat_20042004_132810_16601/app_query.fasta_1.455
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORES=ext -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10078090 @CN 1 1 4237 @runat_20042004_132810_16601 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estm:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	* Match	Query Length	DB ID	Description
1	1261.5	95.1	4569	11	BC014524 Homo sapi
2	1015	76.5	602	10	BE890168 601513147
3	922.5	69.6	651	12	BI732824 603353931
4	896.5	67.6	905	13	BX390092 BX390092
5	860	64.9	729	14	CA321445 UI-M-FWC-
6	686	51.7	693	10	BF969269 602269794
7	662	49.9	904	14	CB209310 AGENCOURT
8	593.5	44.8	643	14	CD310081
c	581	43.8	1160	14	CF661317 CCLM09834
9	564	42.5	396	14	CD773455 AMGNNUC:T
10	553.5	41.7	870	14	CD760225 AGENCOURT
11	526	39.7	374	14	CB692138 AMGNNUC:T
12	523	39.4	343	14	CB693399 AMGNNUC:S
13	519	39.1	383	10	BB872069
14	512	38.6	325	10	BF953476 RC3-NN118
15	509	38.4	324	10	BF953475 RC3-NN118
16	504	38.0	342	10	BF953545 RC3-NN118
17	497	37.5	343	10	BF953479 RC3-NN118
18	488	36.8	323	10	BF953542 RC3-NN118
19	485	36.6	675	14	CA327501 UI-M-FVO-
20	482	36.3	308	10	BF953480 RC3-NN118
21	478	36.0	1174	13	BUI66723 AGENCOURT
22	460	34.7	347	10	BF953470 RC3-NN118
23	429	32.4	713	13	EX880758 EX880759
24	429	32.4	736	13	EX885444 EX885444
25	429	32.4	761	13	BX868932 BX868932
26	425	31.7	362	10	BF953546 RC3-NN118
27	421	31.7	310	10	BF953532 RC3-NN118
28	421	31.7	267	14	CA949304
29	401	30.2	267	14	CA949304
30	363.5	27.4	1018	12	BM452626
31	361.5	27.3	334	10	BF953532 RC3-NN118
32	357	26.9	799	12	EG862377 602796268
33	355.5	26.8	534	10	AW742499 up58a09.y
34	345.5	26.4	586	12	BM940881 UI-M-CGOp
35	334	25.2	531	10	BF074654 222121 MA
36	319	24.1	641	28	AZ625547 1M0465812
37	319	24.1	641	28	AZ625547 1M0465812
38	302	22.8	1139	13	BX344632 BX344632
39	297.5	22.4	668	13	BU610146 UI-M-DJ1-
40	285.5	21.5	1201	9	AL556078 AL556078
c	280.5	19.6	888	29	AL270155 Tetraodon
c	260.5	19.6	961	29	AL272891 Tetraodon
c	260	19.6	305	9	AA938876 op74b06.s
43	259.5	19.6	1045	29	CNS04F1L
44	234	17.6	907	13	BQ433907
45	234	17.6	907	13	BQ433907 AGENCOURT

ALIGNMENTS

RESULT 1
BC014524
LOCUS
DEFINITION
IMAGE:3834205), with apparent retained intron.
ACCESSION
BC014524
VERSION
BC014524.1
KEYWORDS
HTC.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 4569)

AUTHORS

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T.I., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahy, J., Helton, E., Kettman, M., Madan, A., Rodrigues, S., Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalek, U., Smalios, D.E., Schnerch, A., Schein, E., Jones, S.J., and Marra, M.A. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22398257
12477932
2 (bases 1 to 4569)
Strausberg, R.
Direct Submission
Submitted (24-SEP-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DTF
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
Contact: amadams@systemsbiology.org
Anup Madan, Jessica Fane, Erin Helton, Mark Kerteman, Anuradha Madan, Stephanie Rodriguez, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 14 Row: f Column: 19
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis
This clone has the following problem: retained intron.

FEATURES

Location/Qualifiers
1..4569
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3834205"
/tissue_type="Ovary, adenocarcinoma"
/clone_lib="NIH_MGC_9"
/lab_host="DH10B-R"
/note="Vector: pOTB7"

ORIGIN

Alignment Scores:
Pred. No.: 1,49e-101 Length: 4569
Score: 1261.50 Matches: 250
Percent Similarity: 96.17% Conservative: 1
Best Local Similarity: 95.79% Mismatches: 9
Query Match: 95.14% Indels: 1
DB: 11 Gaps: 1

US-10-078-090-151 (1-260) x BC014524 (1-4569)

1 LeuArgAlaLysAlaHisGluGlyLeuGlyPheSerLeuArgGlySerGluHis 20

1582 TTGCGGCGTGCACAGGCCACAGAGGGCTTGGGCTTCAGCATCCGTGGGGCTCGAGCAC 1641
21 GlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluGlySerGlyLeu 40
1642 GCGTGGGATCTACGTGCTCTGTGTGAACACAGGCTCTTACGTGAGAAGAGAGACTG 1701
41 ArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAla 60
1702 CCGGTTCGGGACACAGATTCTGCGCGTCAACAGCAATCTCTGCGCGGTGACCCACGCG 1761
61 GluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAlaGly 80
1762 GAGCGCGTCAAGGCTCTGAAGGGCTCCAAAGAGTGGTGTCTGTCTGTACTCAGCAGG 1821
81 ArgIleProGlyTyrValThrAsnHisIleTyrThrTyrValAspProGlnGlyArg 100
1822 CGCATCCCTGGGGCTACGTCAACACCATCTACACCTGGGTGGACCCGAGGCGCCG 1881
101 SerIleSerProSerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlu 120
1882 AGCATCTCCCCACCTCGGGCTGCCCGCCACCGCGGTGGTCCCTGAGGCAGCAGGAG 1941
121 GlyAspArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysLysValAsnLeu 140
1942 GGTGACCGGAGGAGCACCTCTGCACCTCTGCAAGGAGGGGATGAGAAAAGGTGAACCTG 2001
141 ValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGluTyrGlyLeu 160
2002 GTGCTGGGAGCGCGCTCTCTGGGCTCACGATCCGTGGGGAGTGAAGTACGCGCTT 2061
161 GlyIleTyrIleThrGlyValAspProGlySerGluAlaGluGlySerGlyLeuLysVal 180
2062 GGCATTTTACATCACTGGCGTGGACCCAGGCTCTGAAGCAGAGGAGCGCGGCTCAAGTT 2121
181 GlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHisAspGluAla 200
2122 GGGACACAGATTCTAGAAGTGAATGGCGGAGGCTTCTCAACATCTTACACGAGGCT 2181
201 ValArgLeuLysSerSerArgHisLeuIleLeuThrValLysAspValGlyArgLeu 220
2182 GTCAGGCTGCTTAAGTCACTCGGCACCTCATCTCGACAGTGAAGGAGCTCGGAGGCTG 2241
221 ProHisAlaArgThrValAspGluThrIleValSerPheLeuSerArgIleArgGlu 240
2242 CCCATGCGCGCACCATCTGGGACGAGACCAAGTGGATCCCGAGTCCCGGATCAGGGAG 2301
241 ThrMetAlaAsnSerAlaGly---SerGlyHisSerAlaArgSerAsnLeuThrPro 259
2302 ACCATGGCGAACTCGGAGGCTTTCTTGGCGATCTCACACAGAGGAATAACACAGCCA 2361
260 Gly 260
2362 GGA 2364
RESULT 2
BE890168
LOCUS BE890168
DEFINITION 601513147F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3914675 5', mRNA sequence.
ACCESSION BE890168
VERSION BE890168.1 GI:10348220
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 602)
REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM9736 row: m column: 12
 High quality sequence stop: 602.

FEATURES

Location/Qualifiers
 1..602
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3914675"
 /tissue_type="leiomyosarcoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC 71"
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.1 Kb."

ORIGIN

Alignment Scores:
 Pred. No.: 9,27e-81 Length: 602
 Score: 1015.00 Matches: 200
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 76.55% Indels: 0
 DB: 10 Gaps: 0

US-10-078-090-151 (1-260) x BE890168 (1-602)

QY 21 GlyValGlyTleYrValSerLeuValGluProGlySerLeuAlaGluLysGluGlyLeu 40
 Db 2 GCGGTGGGATCAGCGTCTCTGGTGGACACAGGCTCTAGCTGAGAGAGAGACTG 61
 QY 41 ArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAla 60
 Db 62 CGGTGGGGACAGATTCTCGCGGTCAACGACAAATCCCTGCGCGGTGACCCACCGG 121
 QY 61 GluAlaValLysAlaLeuLysGlySerLysLeuValLeuSerValTyrSerAlaGly 80
 Db 122 GAGGCGCTCAAGGCTCTGAAGGCTCCAGAAAGCTGGTGTCTGTACTCAGCAGGG 181
 QY 81 ArgTleProGlyGlyTyrValThrAsnHisIleTyrThrTyrValAspProGlnGlyArg 100
 Db 182 CGCATCCCTGGGGCTACGTACCAACCAATCTACACCTGGTGGACCGCAGGCGCGC 241
 QY 101 SerIleSerProSerGlyLeuProGlnProHisGlyAlaLeuArgGlnGlnGlu 120
 Db 242 AGCATCTCCACACCTCGGCGCTGCCCCAGCCACCGTGTGCTGAGCGCAGGAG 301
 QY 121 GlyAspArgArgSerThrLeuHisLeuGlnGlyGlyAspGluLysLysValAsnLeu 140
 Db 302 GGTGACCGAGGACACCTTCACCTCTCGAAGGAGGGATGAGAAAGGTGAACCTG 361
 QY 141 ValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGluTyrGlyLeu 160
 Db 362 GTGCTGGGGACGCGCGCTCCCTGGGCTCCAGCATCGTGGGGAGCTGAGTACGCGCTT 421
 QY 161 GlyTleTyrIleThrGlyValAspProGlySerGluAlaGluGlySerGlyLeuLysVal 180
 Db 422 GGCATTTACATCATCTGGCGTGGACCCAGGCTCTGAACGAGGACACCGGCTCAAGGTT 481
 QY 181 GlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnIleLeuHisAspGluAla 200
 Db 482 GGGACACAGATTCTAGAAGTGAATGGGCGGAGCTTTCTCAACATCTCTACACGACGAGCT 541
 QY 201 ValArgLeuLeuLysSerSerArgHisLeuIleLeuThrValLysAspValGlyArgLeu 220
 Db 542 CTCAGGCTGCTTAAGTCATCTCGGCACCTATCTCGACAGTGAAGGACGTCGCGAGGCTG 601

RESULT 3

BI732824
 LOCUS
 DEFINITION
 603353931F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5361582 5', mRNA sequence.
 ACCESSION
 BI732824
 VERSION
 BI732824.1
 KEYWORDS
 EST.
 SOURCE
 Mus musculus (house mouse)
 ORGANISM

REFERENCE
 1 (bases 1 to 651)
 AUTHORS
 NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE
 National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL
 Unpublished (1999)
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: cgabs-rc@mail.nih.gov
 Tissue Procurement: The Cepko Laboratory
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM1919 row: m column: 07
 High quality sequence stop: 651.

FEATURES

Location/Qualifiers
 1..651
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:5361582"
 /tissue_type="retina"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC 94"
 /note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-dt primed. Average insert size 3.3 Kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:
 Pred. No.: 1.96e-72 Length: 651
 Score: 922.50 Matches: 187
 Percent Similarity: 90.57% Conservative: 5
 Best Local Similarity: 88.21% Mismatches: 19
 Query Match: 69.57% Indels: 2
 DB: 12 Gaps: 1

US-10-078-090-151 (1-260) x BI732824 (1-651)

QY 50 AsnAspLysSerLeuAlaArgValThrHisAlaGluAlaValLysAlaLeuLysGlySer 69
 Db 3 AACGATAAATCTTAGCCCGGTGACCCACGCGAGGCTGTCAAGGCTCTCAAGGCTCC 62
 QY 70 LysLysLeuValLeuSerValTyrSerAlaGlyArgIleProGlyGlyTyrValThrAsn 89
 Db 63 AAGAAGCTGGTCTGTCTGTATCTCAGTGGCGGTATCCAGG-CGCTATGTGACCAAC 121
 QY 90 HisIleTyrThrTyrValAspProGlnGlyArgSerIleSerProProSerGlyLeuPro 109
 Db 122 CACATCTACCTGGTGGACCCACAGGCTGAAGCATCCCTCCCTCCAGCCTGCCC 181
 QY 110 GlnProHisGlyGlyAlaLeuArgGlnGlnGluGlyAspArgArgSerThrLeuHisLeu 129
 Db 182 CAGCCCCATGGCAGCACCTTGAGACGGTGAAGATGACCGAAGGAGTACCTCCCTCC 241
 QY 130 LeuGlnGlyGlyAspGluLysLysValAsnLeuValLeuGlyAspGlyArgSerLeuGly 149
 Db 242 CTGCAGAGTGGAGATGAGAAAGGTGAACCTGGTGTGGGGACGCGCGGTCTTGGGC 301
 QY 150 LeuThrIleArgGlyGlyAlaGluTyrGlyLeuGlyIleTyrIleThrGlyValAspPro 169

```

302 CTCACATCCGAGGTGGACAGATGACGCTTGGCAATTATCATCTGGTGTGACCCA 361
170 GlySerGluAlaGluGlySerGlyLeuLysValGlyAspGlnLeuGluValAsnGly 189
362 GGTCTCTGAAGCAGAGACGCGGCTCAAGGTGGAGACCAAGATCTGGAGTGATGG 421
190 ArgSerPheLeuAsnIleLeuHisAspGluAlaValArgLeuLeuLysSerSerArgHis 209
422 CGAGGCTTTCTCAACATCTCTGATGATGAGGAGTGAAGTCTCAAGTCATCCCGGCAC 481
210 LeuIleLeuThrValLysAspValGlyArgLeuProHisAlaArgThrThrValAspGlu 229
482 CTCATCTGACGCTGAAGACGCTGGAGGCTGCCACCGCAGTACCAACGTTGACCAAG 541
230 ThrLysTrpIleAlaSerSerArgIleArgGluThrMetAlaAsnSerAlaGly---Ser 248
542 ACCAAGTGATGCCAGTTCCTCCGATTCGGGGAAGCGTCCCACTCAGCAGGTTTCCA 601
249 GlyHisSerAlaArgSerAsnLeuGlnThrProGly 260
602 GGGGACCACACAGAGAGGACAGCAAGCCAGGA 637

```

```

RESULT 4
BX390092 905 bp mRNA linear EST 08-MAY-2003
LOCUS
DEFINITION
BX390092 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
cDNA clone CS0DK007YH03 5-PRIME, mRNA sequence.
ACCESSION
BX390092
VERSION
BX390092.1 GI:30463208
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

1 (bases 1 to 905)
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 5156.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0BAF042ZH04_AFO3969_1&cluster=5156.r.
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS0BAF042ZH04_AFO3969_1.
Location/Qualifiers
FEATURES
source
1. 905
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DK007YH03"
/cell_type="HELA CELLS COT 25-NORMALIZED"
/call_line="HELA"
/clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

```

```

Alignment Scores:
Pred. No.: 6.67e-70 Length: 905
Score: 896.50 Matches: 186
Percent Similarity: 95.00% Conservativity: 4
Best Local Similarity: 93.00% Mismatches: 6
Query Match: 67.61% Indels: 5
DB: 13 Gaps: 2

```

ORIGIN

```

Alignment Scores:
Pred. No.: 6.67e-70 Length: 905
Score: 896.50 Matches: 186
Percent Similarity: 95.00% Conservativity: 4
Best Local Similarity: 93.00% Mismatches: 6
Query Match: 67.61% Indels: 5
DB: 13 Gaps: 2

```

```

US-10-078-090-151 (1-260) x BX390092 (1-905)
QY 1 LeuArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGlyGlySerGluHis 20
313 TTCCGCGTGCCTCCCAAGCCACACGAGGCTTGGGCTTCAGATCCGTGGGGCTCGAGCAC 372
21 GlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGluGlyLeu 40
373 GCGTGGGCATCTACGTGCTCTGTGGTGAACAGGCTCTCTAGCTGAGAAGGAGACTG 432
41 ArgValGlyAspGlnIleLeuArgValAlaAsnAspLysSerLeuAlaArgValThrHisAla 60
433 CGGCTGGGACACAGATTCTCGCGCTCAACGACAAATCCCTGGCCCGGCTGACCCACGG 492
61 GluAlaValLysAlaLeuLysGlySerLysValLysLeuValLysSerValTyrSerAlaGly 80
493 GAGCCGCTCAAGGCTCTGAAGGCTCCAGAGCTCGTCTGTCTGTCTGTCTGTCTGTCTGT 552
81 ArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTrpValAspProGlnGlyArg 100
553 CGCATCCCTGGGCGTACGTACCAACACCATCTACACCTGGGTGGACCCGCGAGGCCGC 612
101 SerIleSerProSerGlyLeuProGlnProHisGlyValAlaLeuArgGlnGlnGlu 120
613 AGCATCTCCCAACCTCGGCGCTGCCACGCCACCGTGGTCCCTGAGGACGAGGAG 672
121 GlyAspArgArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysLysValAsnLeu 140
673 GGTGACCGGAGGAGCACCTGTCACTTCTGCA-GGAGGGGATGAGAAAAGGTGAACCTG 731
141 ValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGluTyrGlyLeu 160
732 GTGCTGGGACGCGCGCTCCCTGCGCTCACCATCCGTGGGGAGCTGATGACGGCCTT 791
161 GlyIleTyr-IleThrGlyValAspProGlySerGluAlaGluGlySerGlyLeuLysVa 180
792 GGCATTTTACATNCATCTGGCTGGACCCAGCTCTGAAGCAGAGAGGACGGGCTCAGG 851
180 lGlyAspGlnIleLeuGlu---ValAsnGlyArgSerPheLeuAsnIleLeuHisasp 198
852 TGGNGACCAAGTCTAGAGTGAATGGCGGGCTTTCT-----CACATCTCACCGAG 903
CA321445 729 bp mRNA linear EST 09-JUL-2003
UI-M-FW0-cbz-b-04-0-UI.r1 NIH_BMAP_FW0 Mus musculus cDNA clone
IMAGE:6816389 5', mRNA sequence.
CA321445
CA321445.1 GI:24539543
EST.
Mus musculus (house mouse)
SOURCE
Mus musculus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 729)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Straube, Ph.D.
Email: cgbbs-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
The following repetitive elements were found in this cDNA
sequence: 146-183, >GC_rich#low_complexity
seq primer: pyx-5.

```

FEATURES
source

Location/Qualifiers

1. .729
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:5816389"
 /tissue_type="whole brain"
 /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_lib="NIH BMAP FWO"
 /notes="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is AGCGACACG. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

ORIGIN

Alignment Scores:

Pred. No.: 9,03e-67 Length: 729
 Score: 860.00 Matches: 168
 Percent Similarity: 96.02% Conservative: 1
 Best Local Similarity: 95.45% Mismatches: 7
 Query Match: 64.86% Indels: 0
 DB: 14 Gaps: 0

US-10-078-090-151 (1-260) x CA321445 (1-729)

QY 1 LeuArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGlyGlySerGluHis 20
 DB 201 CTGCGGGCGCCAAAGCCCAAGAGGCTGGGGTTCAGCATCCGGGGGCTCGGACAC 260
 QY 21 GlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluGlyLeu 40
 DB 261 GGGTGGGATCATCTACGTGTCTCTAGTGGAGCGGGCTCCCTGGCAGAGGAAGGGTTG 320
 QY 41 ArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAla 60
 DB 321 CGGGTCGGGACCAAGATTCTGCGCGTCAACGATAAATCTTAGCCCGGTGACCCACGCG 380
 QY 61 GluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAlaGly 80
 DB 381 GAGGCTGTCAAGGCTCTCAAGGCTCCAAAGAGCTGGTGTCTGTATACATCAGCTGGG 440
 QY 81 ArgIleProGlyTyrValThrAsnHisIleTyrThrTyrValAspProGlnGlyArg 100
 DB 441 CGTATCCAGGGGGCTATGTGACCAACCACTACCTCTGGGTGACCCACAGGGTGA 500
 QY 101 SerIleSerProSerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlnGlu 120
 DB 501 AGACATATCCCTCCCTCCAGCTCCGCCAGCCGCCATCGGAGCACCCTGAGACAGCGTGA 560
 QY 121 GlyAspArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysValLeuLeu 140
 DB 561 GATGACGAAGAGTACCTTCCCTCCCTGAGAGTGGAGATGAGAAAAGGTGAACCTG 620
 QY 141 ValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGluTyrGlyLeu 160
 DB 621 GTGTTGGGGACGGCGGCTCTTGGGCTCCAGATCCGAGGTGGAGCAGATACGGCCCT 680
 QY 161 GlyIleTyrIleThrGlyValAspProGlySerGluAlaGluGlySer 176
 DB 681 GGCATTTACATCACTGCTGTGGAGCCCGGCTCTGAAGCAGAGAGCAGC 728

RESULT 6

BF969269
 LOCUS 602269794F1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4358158 5', linear
 DEFINITION mRNA sequence.
 ACCESSION BF969269
 VERSION BF969269.1 GI:12336484
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 693)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs@mail.nih.gov

Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM9995 row: k column: 23
 High quality sequence stop: 643.

FEATURES
source

1. .693
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4358158"
 /tissue_type="adrenal cortex carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_84"
 /notes="Organ: adrenal gland; Vector: pCMV-SPOrt6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.229 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:
 Pred. No.: 3.11e-51 Length: 693
 Score: 686.00 Matches: 140
 Percent Similarity: 95.95% Conservative: 2
 Best Local Similarity: 94.59% Mismatches: 6
 Query Match: 51.73% Indels: 1
 DB: 10 Gaps: 0

US-10-078-090-151 (1-260) x BF969269 (1-693)

QY 1 LeuArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGlyGlySerGluHis 20
 DB 36 TTGCGGGTGCACAGGCCCAAGGGCTTGGGCTTCAGCATCCGTGGGGGCTCGAGCAC 95
 QY 21 GlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGluGlyLeu 40
 DB 96 GGGCTGGGATCTACGTGTCTCTGGTGGACACAGGCTCTTAGCTGAGAGGAAGGACTG 155
 QY 41 ArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAla 60
 DB 156 CGGTTGGGGACCAAGATTCTGCGGCTCAACGACAAATCCCTGGCCCGGGTGACCCACGCG 215
 QY 61 GluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAlaGly 80
 DB 216 GAGGCGGTCAAGGCTCTGAAGGGCTCCAAAGAGCTGGTGTCTGTGTACTACGACAGG 275
 QY 81 ArgIleProGlyTyrValThrAsnHisIleTyrThrTyrValAspProGlnGlyArg 100
 DB 276 GGCATTTACATCACTGCTGTGGAGCCCGGCTCTGAAGCAGAGAGCAGC -GGCGCG 334


```
/db_xref="taxon:7668"
/clone="WPMGP69ID1338;MPI_SURUDI_38D13"
/tissue_type="whole larva"
/dev_stage="larva 2-3 weeks"
/lab_host="E.coli, X11 blue"
/clone_lib="Sea urchin larva cDNA library MPMGP691"
/note="Vector: pSport1; Site 1: NotI; Site 2: SalI; Random
primed and directionally cloned in pSport1_vector using a
NotI (5'-TGACCTAGTCTAGTACGAGCGGCC (1)15-3' and a
SalI 5'-TGACCCACGCGTCCG-3' adapters (Gibco BRL)"

ORIGIN
Alignment Scores:
Pred. No.: 5,29e-43 Length: 643
Score: 593.50 Matches: 119
Percent Similarity: 75.0% Conservative: 40
Best Local Similarity: 56.13% Mismatches: 50
Query Match: 44.76% Indels: 3
DB: 14 Gaps: 1

US-10-078-090-151 (1-260) x CD310081 (1-643)
Qy 16 GlyGlySerGluHisGlyValGlyIleThrValSerLeuValGluProGlySerLeu--A 35
Db 14 GTGGGGGGGACATTCGGTGGGATCTTGTGTCGCTGGTGGACCAACAGTTATGGG 73
Qy 35 laGluysGluGlyLeuArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuA 55
Db 74 CAGATAAAGAGGGTTAATCAAGAGGACCATGATTCAGGTCATCATTCATTG 133
Qy 55 laArgValThrHisalaGluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuS 75
Db 134 AGAAGGTAGCCCACTCTGATGTCAGTGAAGATCTTAAGGCGGTCAACAGTGGTCTTGT 193
Qy 75 exValTySerAlaGlyArgIleProGlyGlyTyThrValThrAsnHisIleTyThrtpv 95
Db 194 ATGTCAAGAGTGTGGCGTGTCCCTGCTCTCTCTCTCCACAGACCTACATGGG 253
Qy 95 alAspProGlnGlyArgSerIleSerProProSerGlyLeuProGlnProHisGlyGlyA 115
Db 254 TGAACCCCAAGGGGGGAGTGTGTCCTCCACCGAGGTC---GACCCCTTGTGGGA 310
Qy 115 laLeuArgGlnGlnGluGlyAspArgArgSerThrLeuHisLeuLeuGlnGlyGlyAsp 135
Db 311 GGATGCTTAATGATACACAGATAGAGAGTGGTCTCAACCTCTCAAGAACGCGGATG 370
Qy 135 luLysLysValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyG 155
Db 371 AAAAGAAGTCAATGTGTAGTGAATGAAGGAGAGAGTCTTGGTTGTATGATTCAGGAG 430
Qy 155 lyAlaGluTyGlyLeuGlyIleTyThrIleThrGlyValAspProGlySerGluAlaGlu 175
Db 431 GAAAGAGGTTGGTCTTGGTATCTTCATCATCGGAATGATTAATCTCTGTGGGAGACC 490
Qy 175 lysSerGlyLeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsnI 195
Db 491 ATGCCAGTCTTAAGGTTGGCGATCAATCTCGATGTCAACTCCAGGAACCTCTCGACA 550
Qy 195 leLeuHisaspGluAlaValArgLeuLeuLysSerSerArgHisLeuLeuLeuThrVal 215
Db 551 TCGAGCACAGAACGAGTGGACCTCTCAATCATCAAGCTGATGATGATGATGATGATCA 610
Qy 215 yeAspValGlyArgLeuProHisAlaArgThr 225
Db 611 AGACGTTGGCAAGTTGCCGTACGCGCGGACG 642

RESULT 9
CF661317/c 1160 bp mRNA linear EST 07-OCT-2003
LOCUS CcLM09a34k21f1 Carp muscle library 1 Cyprinus carpio cDNA clone
DEFINITION 34k21 5', mRNA sequence.
ACCESSION CF661317
VERSION CF661317.1 GI:37558475

KEYWORDS
SOURCE Cyprinus carpio (common carp)
ORGANISM Cyprinus carpio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Cyprinus.
REFERENCE
1 (bases 1 to 1160)
Gracey,A.Y., Fraser,E., Li,W. and Cossins,A.R.
Microarray and EST analysis of the carp (Cyprinus carpio)
transcriptome during environmental stress
Unpublished (2003)
COMMENT
Contact: Andrew R. Cossins
Laboratory for Environmental Gene Regulation
University of Liverpool
School of Biological Sciences, The Biosciences Building, Crown
Street, Liverpool, United Kingdom, L69 7ZS
Tel: +44(0)151-795-4510
Fax: +44(0)151-795-4431
Email: cossins@liv.ac.uk
Vector has been trimmed from this EST.
Plate: 34 row: k column: 21
Seq primer: Triplex 5' LD (5'-CTCGGAGCGGCCATTGTGTGT-3')
High quality sequence start: 35
High quality sequence stop: 917.
FEATURES
Location/Qualifiers
source
1..1160
/organism="Cyprinus carpio"
/mol_type="mRNA"
/db_xref="taxon:7962"
/clone="34k21"
/sex="Male & female"
/tissue_type="Muscle"
/dev_stage="Adult"
/lab_host="E.coli Electromax DH10B"
/clone_lib="Carp muscle library 1"
/note="Vector: pTriplex2; Site 1: Sfil GGCCATTACGGCC;
Site 2: Sfil GGCGCTCGGCC; Serially subcloned cDNA
library prepared from muscle of warm, cold and hypoxia
challenged animals"
ORIGIN
Alignment Scores:
Pred. No.: 1,62e-41 Length: 1160
Score: 581.00 Matches: 120
Percent Similarity: 71.86% Conservative: 23
Best Local Similarity: 60.30% Mismatches: 41
Query Match: 43.82% Indels: 15
DB: 14 Gaps: 2

US-10-078-090-151 (1-260) x CF661317 (1-1160)
Qy 70 LysLysLeuValLeuSerValTySerAlaGlyArg-----IleProGly 84
Db 1024 AAAAGGCTTTAAGAAGGCTCTTTAAGTCGGTCTGTTCAATGGGACGATTCCCGGT 965
Qy 85 GlyTyValThrAsnHisIleTyThrValAspProGlnGlyArgSerIleSerPro 104
Db 964 GGTACGTGACCAATCAGGTTACACATGGTGCACCTCAGGCGCGCAGTGTCTCCCA 905
Qy 105 ProSerGlyLeuProGlnProHis-GlyGlyAlaLeuArgGlnGlnGlyAspArg 124
Db 904 CCACCACTTGTGGGAGCAGCATGGGCTCAGGAGGAAGACATTGGAAATCCAGCACT 845
Qy 124 gSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysValAsnLeuValLeuGlyAs 144
Db 844 TGCCACACCCATTCGGCAGGCTGCTAGAGAGAGGTGAATATCTCATCTGATGA 785
Qy 144 pGlyArgSerLeuGlyLeuThrIleArgGlyAlaGluTyThrGlyLeuGlyIleTy 164
Db 784 TGGCCATCTCTGGGTTTGATGATCCGTGAGGTGCTGATGATGCTCTGGGCATTATAT 725
Qy 164 eThrGlyValAspProGlySerGluAlaGluGlySerGlyLeuLysValGlyAspGln 184
```



```

Db 724 CACCGAGTGGACCGAGGCTCTCGCAGAGTACACGGGACTAAAGGTAGGGATCAGAT 665
Qy 184 eLeuGluValAsnGlyArgSerPheLeuAsnLeuHisAspGluAlaValArgLeu 204
Db 664 TCTGGAGGTCAATGGAGCTAGCTTCCGGAGCATTTCTCATGACGAGCGCGTGCAGATCTT 605
Qy 204 uLysSerSerArgHisLeuLeuThrValLysAspValGlyArgLeuProHisAla 224
Db 604 GAAGAATCCCGGCACATGCTGATGACCATTAAGGATGTGGGGCGCTCCCTCAGCAGC 545
Qy 224 gThrThrValAspGluThrLysTrpLeuAlaSerArgIleArgGluThrMetAla 244
Db 544 AACTGTGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 485
Qy 244 nSer-----AlaGlySerGlyHisSerAlaArg 253
Db 484 TACCAACACCCAGAGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 430

RESULT 10
CB773455
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
ORIGIN
Alignment Scores:
Pred. No.: 1.15e-40 Length: 396
Score: 564.00 Matches: 118
Percent Similarity: 69.77% Conservative: 2
Best Local Similarity: 68.60% Mismatches: 8
Query Match: 42.53% Indels: 44
DB: 14 Gaps: 2

US-10-078-090-151 (1-260) x CB773455 (1-396)

Qy 55 AlaArgValThrHisAlaGluAlaValLysAlaLeuLysGlySerLysLeuValLeu 74
Db 11 GCCCGGTGACCCACGCGGAGGTGTCAAGGCTCTCAAGGGTCTCAAGAGTGGTGTG 70
Qy 75 SerValTyrSerAlaGlyArgIleProGlyGlyTyrValThrAsnHisIleTyrThr 94
Db 71 TCTGTATACCTGCTGGGCTATACCGGGGGTATGTACCAACACATCTACACCTGG 130
Qy 95 ValAspProGlnGlyArgSerIleSerProProSerGlyLeuProGlnProHisGly 114
Db 131 GTGGACCCCGAGGTGGAAGCACATCCCTCTCCCTCCAGCCTT-----CCCATG 184

```

```

Qy 115 AlaLeuArgGlnGlnGluGlyAspArgSerThrLeuHisLeuLeuGlnGlyGlyAsp 134
Db 185 ACCCTGAGACAGCATGAAGATGATCGAAGAGTGCCTACACCTCTCGCAGAGTGGAGAT 244
Qy 135 GlulysLysValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGly 154
Db 245 GAGAAA----- 250
Qy 155 GlyAlaGluTyrGlyLeuGlyIleTyrIleThrGlyValAspProGlySerGluAlaGlu 174
Db 250 ----- 250
Qy 175 GlySerGlyLeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsn 194
Db 251 -----AAGTTGGGACCATGTTCTCGAGTGAACGGCGGAGCTTCTCAGC 298
Qy 195 IleLeuHisAspGluAlaValArgLeuLysSerSerArgHisLeuLeuThrVal 214
Db 299 ATCTGTCATGACGAGGAGTGAAGCTGCTCAAGTCATCCCGGACCTCATCTCTGACGCGTG 358
Qy 215 LysAspValGlyArgLeuProHisAlaArgThr 226
Db 359 AAGACGTCGGAGGCTGCCCCACGACGACGACCC 394

RESULT 11
CD760225
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
ORIGIN
Alignment Scores:

```

```

/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="trp2-00004-d6"
/tissue_type="peneal gland brain"
/clone_lib="trp2 (10294)"
/notes="Vector: C6KGF7L; Site_1: SalI; Site_2: NotI;
peneal gland brain region"

ORIGIN
Alignment Scores: 2.67e-37 Length: 374
Pred. No.: 526.00 Matches: 111
Score: 68.48% Conservative: 2
Percent Similarity: 67.27% Mismatches: 8
Best Local Similarity: 39.67% Indels: 44
Query Match: 14 Gaps: 2
DB:

US-10-078-090-151 (1-260) x CB692138 (1-374)
QY 55 AlaArgValThrHisAlaGluAlaValLysAlaLeuLysGlySerLysLysLeuValLeu 74
Db 11 GCCCGGTGACCCACGCGGAGGCTGTCAAGGCTCTCAAGGCTCCCAAGAGTTGGTCTG 70
QY 75 SerValTyrSerAlaGlyArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTir 94
Db 71 TCTGTATCTACGTGGCGGTATACCCGGGGCTATGTCAACCAACCATCTACACCTGG 130
QY 95 ValAspProGlnGlyArgSerIleSerProSerGlyLeuProGlnProHisGlyGly 114
Db 131 GTGGACCCCGGAGGTCGAAGCACATCCCTCCCTCCAGCCTT-----CCCCATGGCAGC 184
QY 115 AlaLeuArgGlnGlnGluGlyAspArgSerThrLeuHisLeuLeuGlnGlyGlyAsp 134
Db 185 ACCCTGAGACAGCATGAAGATGATCGAAGAGTGCCCTACACCTCTCTGCAGAGTGAGAT 244
QY 135 GluLysLysValAsnLeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGly 154
Db 245 GAGAAA-----250
QY 155 GlyAlaGluTyrGlyLeuGlyIleTyrIleThrGlyValAspProGlySerGluAlaGlu 174
Db 250 -----250
QY 175 GlySerGlyLeuLysValGlyAspGlnIleLeuGluValAsnGlyArgSerPheLeuAsn 194
Db 251 -----AAGTTGGGGGACACAGATTCTGGAGGTGAACGGCGGAGCTTTCTCAGC 298
QY 195 IleLeuHisAspGluAlaValArgLeuLysSerSerArgHisLeuIleLeuThrVal 214
Db 299 ATCTCTGCATGACGAGCGAGTGAAGCTGTCAAGTCAATCCCGCACCTCATCTGACCGTG 358
QY 215 LysAspValGlyArg 219
Db 359 AAGGACGTCGGGAGG 373

RESULT 13
CB692138
LOCUS
DEFINITION
trp2-00004-d6 5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 374)
Amgen EST Program.
Amgen Rat EST Program
Amgen Rat EST Program
Unpublished (2003)
Contact: Dan Fitzpatrick
Amgen, Inc
One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA
Tel: 805 447-4881
Plate: 00004 row: d column: 6.
Location/Qualifiers
1..374

FEATURES
source

```

```

Pred. No.: 3.1e-39 Length: 870
Score: 553.50 Matches: 118
Percent Similarity: 74.59% Conservative: 17
Best Local Similarity: 65.19% Mismatches: 45
Query Match: 41.74% Indels: 2
DB: 14 Gaps: 1

US-10-078-090-151 (1-260) x CD760225 (1-870)
QY 1 LeuArgArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGlyGlySerGluHis 20
Db 233 CTCAGCGTCACAGAGCAATGAAGTTTGGGATTAGCATCCGCGGAGGATCAGAGCAT 292
QY 21 GlyValGlyTyrValSerLeuValGluProGlySerLeuAlaGluLysGlyGlyLeu 40
Db 293 GGAGTTGGGAATCTATGTTTCTCTGTTGGAACCTGATCACTGCTGGAAGAGAGGTTTG 352
QY 41 ArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAla 60
Db 353 AGAATCGGAGAGCCAAATAATGAAGTTAAGCAAAAGTGTTCACCGGTCACGCGCGC 412
QY 61 GluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAlaGly 80
Db 413 GATCGAGTAAAGGTGTGAAGGACGACAAAAGCTCTGCATGTCGGTGTCTGTGGGC 472
QY 81 ArgIleProGlyGlyTyrValThrAsnHisIleTyrThrTirValAspProGlnGlyArg 100
Db 473 AGGATACCGCGGTTATATACCAATACGCTACACCTGGGTGGATCCTCAGGTCGC 532
QY 101 SerIleSerProSerGlyLeuProGlnProHisGlyGlyAla---LeuArgGlnGln 119
Db 533 AGTGTGTCTCTCCACCGGACCTGCTGGGGAGCATCGCAGTGCGACGAGTGAAGTAAAT 652
QY 120 GluGlyAspArgSerThrLeuHisLeuLeuGlnGlyGlyAspGluLysLysValAsn 139
Db 593 AACAGTCAGCGCGCAGTCACATGCGAGTCTCAGGATGGAGATGAGAAGAAGCTAAAT 652
QY 140 LeuValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGluTyrGly 159
Db 653 CTGGTGTGACGACTGCCCTCTTTGGGTTTGTGATGATCCTCGGATGACTGATTAATT- 711
QY 160 LeuGlyIleTyrIleThrGlyValAspProGlySerGluAlaGluGlySerGlyLeuLys 179
Db 712 CTGGGTATCTATTAAAGGTGAATAGTTTGCAGCGGATTCAGGATGAAGGGATAACT 771
QY 180 Val 180
Db 772 ATC 774

RESULT 12
CB692138
LOCUS
DEFINITION
AMGNNUC:TRPG2-00004-D6-A trp2 (10294) Rattus norvegicus cDNA clone
trp2-00004-d6 5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 374)
Amgen EST Program.
Amgen Rat EST Program
Amgen Rat EST Program
Unpublished (2003)
Contact: Dan Fitzpatrick
Amgen, Inc
One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA
Tel: 805 447-4881
Plate: 00004 row: d column: 6.
Location/Qualifiers
1..374

FEATURES
source

```

COMMENT

Contact: Dan Fitzpatrick
Angen, Inc
One Angen Center Drive, Thousand Oaks, CA 91320-1799, USA
Tel: 805 447-4881
Plate: 00029 row: a column: 8.

FEATURES

Location/Qualifiers
1. .343
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="srp2-00029-a8"
/tissue_type="pituitary gland brain"
/clone_lib="srp2 (10238)"
/note="Vector: pSPOR1; Site_1: SalI; Site_2: NotI; peneal gland brain region"

ORIGIN

Alignment Scores:
Pred. No.: 4,37e-37 Length: 343
Score: 523.00 Matches: 102
Percent Similarity: 90.27% Conservative: 9
Best Local Similarity: 90.27% Mismatches: 0
Query Match: 39.44% Indels: 2
DB: 14 Gaps: 1

US-10-078-090-151 (1-260) x CB693359 (1-343)

QY 81 ArgileProGlyGlyTyrValThrAsnHisIleTyrThrTrpValAspProGinglyArg 100
DB 10 CGTATACCGGGGCTATGTACCAACACATCTACACCTGGGTGGACCCCGAGGTGCA 69
QY 101 SerIleSerProSerGlyLeuProGlnProHisGlyGlyAlaLeuArgGlnGlu 120
DB 70 AGCATATCCCTCCCTCCAGCCTT-----CCCATGGCAGCACCTTGACACATGAA 123
QY 121 GlyAspArgSerThrLeuHisLeuGlnGlyAspGlyLysValAsnLeu 140
DB 124 GATGATCGAGGAGTGCCCTACACCTCTCGAGAGTGGAGATGAGAAAGGTGACCTG 183
QY 141 ValLeuGlyAspGlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGluTyrGlyLeu 160
DB 184 GTGCTGGGGGACGGCGGCTCTCTGGGCTCAATCCGAGGTGGAGCGGAGTACGGCCTT 243
QY 161 GlyIleTyrIleThrGlyValAspProGlySerGluAlaGluGlySerGlyLeuVal 180
DB 244 GCGATTATACATCACTGGTGTGGACCGGGCTCTGAAGCCGAGCAGCGGCTCAGGTT 303
QY 181 GlyAspGlnIleLeuGluValAsnGlyArgSerPheLeu 193
DB 304 GGGGACCAAGATTCTGGAGGTGACGGCGGAGCTTTCTC 342

RESULT 14

BB872069 383 bp mRNA linear EST 27-NOV-2001
LOCUS BB872069
DEFINITION BB872069 RIKEN full-length enriched, 1 month neonate cerebellum Mus musculus cDNA clone G630034G23 5', mRNA sequence.

ACCESSION BB872069
VERSION BB872069.1 GI:17119279

KEYWORDS

SOURCE Mus musculus (house mouse)

ORGANISM

Mus musculus
Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.
1 (bases 1 to 383)
Akamatsu, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Imotani, K., Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Soabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toy, T., Watabiki, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

TITLE

JOURNAL
COMMENT

RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al. 2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Wakeda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001).
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

e mouse tissues.

Location/Qualifiers
1. .383
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="G630034G23"
/tissue_type="cerebellum"
/dev_stage="1 month neonate"
/clone_lib="RIKEN full-length enriched, 1 month neonate cerebellum"

ORIGIN

Alignment Scores:
Pred. No.: 1.17e-36 Length: 383
Score: 519.00 Matches: 103
Percent Similarity: 97.17% Conservative: 0
Best Local Similarity: 97.17% Mismatches: 3
Query Match: 39.14% Indels: 0
DB: 10 Gaps: 0

US-10-078-090-151 (1-260) x BB872069 (1-383)

QY 1 LeuArgArgAlaLysAlaHisGluGlyLeuGlyPheSerIleArgGlyCysGluHis 20
DB 66 CTGCGCGCGCCAGGCCACGACGCGGCTTGGGCTTCAGCATCGGGGGGCTCGGAACAG 125
QY 21 GlyValGlyIleTyrValSerLeuValGluProGlySerLeuAlaGluLysGluGlyLeu 40
DB 126 GGGGTGGGCATACCGTGTCTTAGTGGACCGGGCTCCCTGGCAGAGAAGAGGGTTG 185
QY 41 ArgValGlyAspGlnIleLeuArgValAsnAspLysSerLeuAlaArgValThrHisAla 60
DB 186 CGGTCGGGACGAGATTCGCGCGTCAACGATAATCTCTAGCCGGGTGACCCACGG 245
QY 61 GluAlaValLysAlaLeuLysGlySerLysLysLeuValLeuSerValTyrSerAlaGly 80
DB 246 GAGGCTCTCAAGGCTCTCAAGGCTCAAGAGGCTCGTCTGTCTGTACTCAGCTGG 305
QY 81 ArgIleProGlyTyrValThrAsnHisIleTyrThrTrpValAspProGinglyArg 100

Db 306 CGTATCCCGGGGGTGTGTGACCAACCAAAATCTTACACCTGGGTGGACCCACAGGTCGA 365
 QY 101 SerIleSerProProSer 106
 Db 366 AGCACATCCCTCCCTCC 383

RESULT 15

BF953476/c
 LOCUS RC3-NN1182-121100-011-g08 NN1182 Homo sapiens linear EST 22-JAN-2001
 DEFINITION RC3-NN1182-121100-011-g08 NN1182 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BF953476
 VERSION BF953476.1 GI:12370751
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 325)
 AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bai, G.S., Simpson, D.H.,
 Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V.,
 O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

TITLE

sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?ti=RC3&t2=RC3-NN1182-121100-011-908&t3=2000-11-12&t4=1>)
 Seq primer: puc 18 forward
 High quality sequence stop: 325.

FEATURES

source

1..325
 /location/qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /dev_stage="Adult"
 /clone_lib="NN1182"
 /notes="Organ: nervous normal; Vector: puc18; Site_1: SmaI;
 Site_2: SmaI; A mini-library was made by cloning products
 derived from ORBESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the pUC 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."

ORIGIN

Alignment Scores:
 Pred. No.: 3.9e-36 Length: 325
 Score: 512.00 Matches: 102
 Percent Similarity: 99.03% Conservative: 0
 Best Local Similarity: 99.03% Mismatches: 1
 Query Match: 38.61% Indels: 0
 DB: 10 Gaps: 0

US-10-078-090-151 (1-260) x BF953476 (1-325)

QY 125 SerThrLeuHisLeuLeuGlyGlyAspGlyLysValAenLeuValLeuGlyAsp 144
 Db 320 AGCACCTGCTCCCTCCCTCCAGGAGGGGATGAGAAAAGGTGAACCTGGTGTGGGGAC 261

QY 145 GlyArgSerLeuGlyLeuThrIleArgGlyGlyAlaGluTyrGlyLeuGlyIleTyrIle 164
 Db 260 GGCCGGTCCCTGGGCTCAGATCCGTGGGGAGCTGAGTACGGCCTTGGCATTTACATC 201
 QY 165 ThrGlyValAspProGlySerGluAlaGluGlySerGlyLeuLysValGlyAspGlnIle 184
 Db 200 ACTGGCGTGGACCCAGGCTCTGAAGCAGAAAGGCAGCGGGCTCAAGGTTGGGGACCAATT 141
 QY 185 LeuGluValAsnGlyArgSerPheLeuAsnIleLeuHisAspGluAlaValArgLeuLeu 204
 Db 140 CTAGAAAGTGAATGGCGGAGCTTTCTCAACATCCTTACACGACGAGGCTGTCAGGCTGCTT 81
 QY 205 LysSerSerArgHisLeuIleLeuThrValLysAspValGlyArgLeuProHisAlaArg 224
 Db 80 AAGTCATCTCGGCACCTCATCTGACAGTGAAGGACGTGGGAGGCTGCCCATGCCCGC 21
 QY 225 ThrThrVal 227
 Db 20 ACCATTGTG 12

Search completed: April 22, 2004, 03:41:32
 Job time : 2521 secs